

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

На правах рукопису
УДК 519.2

До захисту допущено
В. о. завідувача кафедри ММСА
О.Л.Тимошук
«__» _____ 2019 р.

Магістерська дисертація

на здобуття ступеня магістра за спеціальністю 124 Системний аналіз
на тему: «Моделювання нестационарних процесів ціноутворення»

Виконав:

студент II курсу, групи КА-81 мп
Чупрін Денис Станіславович

Керівник: доцент кафедри ММСА,
д.т.н, проф. Бідюк П.І.

Рецензент: професор кафедри
інформаційної безпеки
КПІ ім. Ігоря Сікорського
д.т.н., проф. Архипов О.Є.

Засвідчую, що у цій магістерській дисертації
немає запозичень з праць інших авторів
без відповідних посилань
Студент _____

Київ
2019

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
 «КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
 ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
 ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ
 КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

Рівень вищої освіти — другий (магістерський)

Спеціальність (спеціалізація) — 124 «Системний аналіз» («Системний аналіз і управління»)

ЗАТВЕРДЖУЮ

В. о. завідувача кафедри ММСА

О.Л.Тимощук

«___» _____ 2019 р.

ЗАВДАННЯ

на магістерську дисертацію студенту Чупріну Денису Станіславовичу

1. Тема дисертації: «Моделювання нестационарних процесів ціноутворення», науковий керівник дисертації Бідюк Петро Іванович, к.т.н., доцент, затверджені наказом по університету від 08.11.2019 р. № 3862-с
2. Термін подання студентом дисертації: 13.12.2019 р.
3. Об'єкт дослідження: нестационарні процеси у ціноутворенні.
4. Предмет дослідження: статистичні та регресійні методи моделювання нестационарних процесів; статистичні критерії для аналізу адекватності моделей і якості оцінок прогнозу
5. Перелік завдань, які потрібно розробити:
 - 1) виконати аналіз методів математичного моделювання нестационарних процесів ціноутворення;
 - 2) дослідити статистичні тести для аналізу даних на наявність тренду і гетероскедастичності;
 - 3) побудувати адекватні моделі вибраних процесів ціноутворення;
 - 4) обчислити оцінки короткострокових прогнозів досліджених процесів та проаналізувати їх якість;
 - 5) сформулювати перспективи подальших досліджень.
6. Орієнтовний перелік графічного (ілюстративного) матеріалу:

- 1). Побудова моделі. Схема структурна
 - 2). Результати статистичних тестів
 - 3). Результати аналізу на інтегрованість, коінтегрованість, гетероскедастичність.
 - 4). Корелограма. Результати
 - 5). Аналіз адекватності моделі. Результати.
 - 6). Обчислення прогнозу. Графік, чисельні данні
 - 7). Аналіз якості прогнозу. Результати.
8. Дата видачі завдання: 05.09.2019 р.

КАЛЕНДАРНИЙ ПЛАН

№ зп	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації
1	Розробка структури, вступу	05.09.19—08.09.19
2	Розробка списку літератури, оформлення	09.09.19—25.09.19
3	Виконання 1-го розділу	26.09.19—05.10.19
4	Виконання 2-го розділу	05.10.19—12.10.19
5	Підбір тестових даних	13.10.19—24.10.19
6	Побудова і аналіз моделі тест даних	25.10.19—02.11.19
7	Підбір даних до основної моделі	02.11.19—05.11.19
8	Побудова і аналіз основної моделі	06.11.19—12.11.19
9	Формулювання результатів дослідження та висновків	12.11.19—20.11.19

Студент

Д.С.Чупрін

Науковий керівник

П.І.Бідюк

РЕФЕРАТ

Магістерська дисертація: 99 стор., 19 іл., 28 табл., 13 бібл. найм.

Актуальність теми. Сучасні методи аналізу фінансових процесів значною мірою ґрунтуються на використанні математичних моделей, які описують динаміку самого процесу. Модулі використовуються для розв'язання задачі короткострокового прогнозування.

Мета та задачі дослідження. Побудова адекватних моделей нестационарних процесів ціноутворення для оцінювання короткострокових прогнозів досліджуваних процесів.

Рішення поставлених завдань та досягнуті результати. Виконано аналіз методів мат моделювання нестационарних процесів ціноутворення . Досліджено статистичні тести для аналізу даних на наявність тренду і гетероскедастичності

Об'єкт досліджень: нестационарні процеси у ціноутворенні.

Предмет досліджень: математичні моделі досліджуваних процесів статистичні критерії для аналізу адекватності моделей і якості оцінок прогнозу.

Методи досліджень: регресійний аналіз; статистичний аналіз даних даних за допомогою спеціальних тестів.

Наукова новизна: нові математичні моделі вибраних процесів ціноутворення; удосконалення методики регресійного моделювання фінансових процесів; оцінки короткострокових прогнозів.

Практичне значення одержаних результатів. Отримані результати можуть використовуватись для виконання математичних обчислень людьми, не знайомими із програмуванням.

ІНТЕГРОВАНІСТЬ, КОІНТЕГРОВАНІСТЬ, ГЕТЕРОСКЕДАСТИЧНІСТЬ,
АДЕКВАТНІСТЬ МОДЕЛІ, ТРЕНДИ

ABSTRACT ON MASTER'S THESIS

The master's thesis was done on 99 pages, contains 19 illustrations, 28 tables,
13 sources were used in the preparation

Topicality. Modern methods of analysis of financial processes are largely based on the use of mathematical models that describe the dynamics of the process itself (pricing, profitability, investment processes, price indicators, etc.), trend, variance, periodic effects, abrupt changes and various nonlinear effects.

Purpose. Building adequate models of non-stationary pricing processes to evaluate the short-term forecasts of the processes under study.

Solution. The analysis of methods of mat modeling of non-stationary pricing processes is performed. Statistical tests are investigated to analyze data for trend and heteroskedasticity.

Object of research. Non-stationary processes in pricing

Subject of research . Statistical criteria for analyzing the adequacy of models and the quality of forecast estimates

Research methods: regression analysis; statistical analysis of data by special tests.

Scientific novelty: new mathematical models of vibration pricing processes; improvement of regression modeling methods of different processes; Short-term forecasts.

The practical value of the results: The results obtained can be used to perform mathematical calculations by people unfamiliar with programming, including mathematical equations and so on.

INTEGRITY, COINTEGRATION, HETEROSKEDASTICITY,
MODEL ADEQUACY, TRENDS

З М І С Т

ВСТУП	8
РОЗДІЛ 1. ЗАДАЧІ МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ СТАЦІОНАРНИХ І НЕСТАЦІОНАРНИХ ПРОЦЕСІВ	10
1.1. Існуючі системи для моделювання і прогнозування стаціонарних і нестаціонарних процесів	10
1.1.1. Графічний аналіз і основні описові статистики	11
1.2.2. Регресійний аналіз	12
1.1.3. Нелінійна регресія	12
1.1.4. Моделі із дискретною залежною змінною	12
1.1.5. Ідентифікація моделі	13
1.1.6. Моделювання нестаціонарних рядів	14
1.2. Типи нестаціонарних процесів	14
1.2.1. Інтегровані процеси	16
1.2.2. Гетероскедастичні процеси	17
Висновки до розділу 1	19
РОЗДІЛ 2. МЕТОДИ МОДЕЛЮВАННЯ І ПРОГНОЗУВАННЯ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ, КРИТЕРІЇ АДЕКВАТНОСТІ	21
2.1. Моделі процесів з детермінованим трендом	21
2.1.1. Тест на тренд	22
2.2. Поліноміальна регресія	23
2.3. Моделі випадкових трендів	24
2.4. Моделі гетероскедастичних процесів	33
2.5. Моделі стаціонарних процесів	40
2.5.1 Модель авторегресії	40
2.5.2 Модель ковзного середнього	42
Висновки до розділу 2	47
РОЗДІЛ 3. ПОБУДОВА МОДЕЛЕЙ ТА ОЦІНЮВАННЯ ПРОГНОЗІВ	48
3.1. Архітектура процесу побудови	48

3.2. Перевірка присутності нестационарності (тест Дікі-Фуллера)	49
3.3. Розширений тест Дікі-Фуллера	53
3.4. Перевірка процесу на стаціонарність (наявність тренду) за допомогою тесту Дікі-Фуллера	54
3.5. Критерії адекватності моделі	59
3.6. Побудова моделей вибраних процесів за статистичними даними	63
3.7. Дослідження та моделювання ряду Лоренца	68
3.8. Дослідження та моделювання ВВП України	73
3.9. Дослідження та моделювання ВВП США	77
РОЗДІЛ 4. СТАРТАП-АНАЛІЗ ПРОЕКТУ	82
4.1. Інформаційна карта проекту	82
4.2. Технологічний аудит ідеї проекту	83
4.3. Аналіз ринкових можливостей запуску стартап-проекту	85
4.4. Розроблення ринкової стратегії проекту	91
Висновки до розділу 4	96
ВИСНОВКИ	97
ПЕРЕЛІК ПОСИЛАНЬ	98

ВСТУП

Майбутнє багатьох явищ невідоме, але воно дуже важливе для рішень, які приймаються в даний момент. Прогнозування – найважливіший компонент аналітичної роботи, що дозволяє передбачити найбільш ймовірний розвиток подій, а також визначити, які заходи впливу приведуть до тих чи інших результатів.

Прогнозування – один з вирішальних елементів ефективної організації управління окремими господарюючими суб'єктами та економічними спільнотами внаслідок того, що якість прийнятих рішень великою мірою визначається якістю прогнозування їх наслідків. Тому рішення, прийняті сьогодні, повинні спиратися на достовірні оцінки можливого розвитку досліджуваних явищ та подій в майбутньому.

Прогноз повинен визначати можливості, в рамках яких можуть ставитися реалістичні задачі планування розвитку економіки або роботи підприємства. У прогнозуванні використовуються математичні методи, часом досить складні і різні для конкретних завдань. Поширення комп'ютерів, простих та ефективних програмних продуктів робить прогнозування більш швидким і досить ефективним механізмом аналізу.

У роботі проведено огляд існуючих на даний момент продуктів для статистичної обробки даних та аналізу природи нестационарних процесів і їх класифікації. Розглянуто типи архітектури, які використовуються для проектування СППР. Також описано поширені методи оцінювання параметрів моделі та прогнозування для стаціонарних і нестационарних процесів.

На основі моделі АРКС реалізована СППР, із допомогою якої можливо визначати найкращі моделі з мінімальними на це затратами ресурсів. Також передбачено гнучкий інтерфейс для роботи з програмним продуктом.

Перший розділ присвячено аналізу існуючих на даний момент продуктів для статистичної обробки даних, аналізу природи нестационарних процесів і їх

класифікації. Розглянуто типи архітектур, які використовуються для проектування СППР. Другий розділ описує поширені методи прогнозування для стаціонарних і нестаціонарних процесів. Також в цьому розділі розглянуто методи оцінювання параметрів моделі. В третьому розділі описано спроектовану СППР, продемонстровано основні її характеристики та можливості. В четвертому розділі розглянуті приклади роботи системи на реальних фінансово-економічних процесах. П'ятий розділ стосується фінансовартісного аналізу створеного програмного продукту. В шостому розділі наведена характеристика умов праці згідно існуючих нормативів, а також приведені рекомендації щодо їх покращення.

РОЗДІЛ 1. ЗАДАЧІ МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ СТАЦІОНАРНИХ І НЕСТАЦІОНАРНИХ ПРОЦЕСІВ

1.1. Існуючі системи для моделювання і прогнозування стаціонарних і нестаціонарних процесів

Для опису динаміки процесів (доходність, ціноутворення) використовують математичні моделі, що є ґрунтом для сучасних методів аналізу фінансових процесів.

Динаміка дисперсії процесу (або відповідного стандартного відхилення) потрібна для аналізу стаціонарності, фінансових ризиків, виконання торгових операцій і формування портфелів фінансових активів, побудови СППР для керування фінансовими потоками, а тому часто її необхідно моделювати. В аналізі фінансових процесів стандартне відхилення і є волатильністю. Візьмемо випадкову складову часового ряду, що описує поведінку фінансової змінної. Волатильність можна визначити як міру її мінливості.

Так як волатильність є неспостережуваним параметром, оцінювати її необхідно на основі спостережень основної змінної. Для щоденних значень фінансового процесу (доходності) використовуються квадрати інновацій. Це є простішим способом оцінки волатильності при аналізі можливих втрат. Квадрати інновацій (зашумлені історичні оцінки волатильності) відображають основну властивість волатильності, а саме наявність змін у часі. У волатильності спостерігається персистентність, так як поточні її значення пов'язані з минулими. Вплив відповідних цінових шоків на величину ризику та їх тривалість визначається за допомогою ступеня цього взаємозв'язку. Наявність персистентності забезпечують більш точне прогнозування волатильності.

Узагальнена авторегресія із умовною гетероскедастичністю є ключовою моделлю для опису умовної дисперсії і містить структуру авторегресії з

ковзним середнім [2, 3]. Але модель містить недоліки, адже згідно з її описом, волатильність експоненційно затухає в часі, що не завжди співпадає з дійсністю. «Довга пам'ять» передбачає спадання з меншою швидкістю, а саме гіперболічне спадання цінових шоків, а тому краще описує волатильність. Але слід зазначити, що волатильність не завжди має такий тип поведінки.

За допомогою встановлення структурних змін у відповідних часових рядах, ми, досліджуючи фінансові процеси, маємо змогу відстежувати довгострокові ефекти, в тому числі і волатильність.

1.1.1. Графічний аналіз і основні описові статистики

Побудова різних графіків, діаграм, корелограм, спектрограм потрібна для легкого візуалізування кінцевих результатів обчислень або вихідних параметрів моделі. За допомогою команд меню здійснюється перегляд вихідних даних в графічній інтерпретації, за допомогою гістограм, лінійних графіків і так далі.

За допомогою відповідних кнопок з робочого модуля виробляється графічний перегляд результатів в програмі Statistica. Після кожної завершеної процедури, додаток автоматично дає змогу задати налаштування побудови графіка, також можна переглядати данні в різних масштабах. Різницею між Statistica та Eviews в побудові АКФ і ЧАКФ є виконання їх побудови в різних вікнах у Statistica і в одному вікні у Eviews. Використовуючи команди основного меню робочого файлу можна переглянути графічні результати моделювання в Eviews. Для перегляду числових характеристик тестових даних в Eviews використовуються команди меню, в Statistica для цього передбачений окремий модуль. Для обох пакетів набір описових статистик є однаковим.

1.2.2. Регресійний аналіз

Регресійний аналіз — розділ математичної статистики, присвячений методам аналізу залежності однієї величини від іншої. Займається пошуком моделі зв'язку, вираженої у функції регресії. [2]

Якщо відношення між змінними можуть бути виражені через комбінацію цих змінних, використовується регресійний аналіз. Отримана комбінація застосовується для прогнозування значення, що приймає залежна(цільова змінна), що отримується завдяки набору значень незалежних(вхідних) змінних. При найпростіших випадках, для цього використовується лінійна регресія. Але велика кількість реальних моделей є поза рамками лінійної регресії. Це породжує необхідність комплексних методів, що забезпечать прогнозування майбутніх значень.

1.1.3. Нелінійна регресія

Для оцінювання коефіцієнтів нелінійної регресії довільного вигляду використовуються ітеративні методи. Прикладами деяких таких методів є : квазі-ньютонівський, Хука-Джівса, симплексний. В Eviews застосовуючи тест Боксу-коксу можна підібрати нелінійну функцію, що матиме найкраще відображення залежності. Використовуючи нелінійний МНК (NLS) і зважений МНК (WLS) здійснюється оцінка коефіцієнтів.

1.1.4. Моделі із дискретною залежною змінною

Вихідними даними для нелінійного оцінювання є : R^2 -статистика , логарифмічна та обмежена логарифмічна функція правдо-подібності та інші.

Використовуючи Eviews можна будувати моделі бінарного і множинного вибору. Можна використовувати як порядкові, так і неврегульовані альтернативи. Щоб обрати метод, яким буде виконано оцінювання, відведено спеціальне поле, де треба вибрати метод, що відповідає заданій моделі. У якості вхідних даних можуть виступати R2-статистика, псевдо-коефіцієнт детермінації, логарифмічні функції правдоподібності.

1.1.5. Ідентифікація моделі

Ідентифікація моделі АРКС так проводиться за допомогою q-статистики. Також, по критеріям Шварца і Акайке, можна зробити селекцію двох моделей, тобто порівняти їх. Інструментом виконання цих критеріїв в Eviews виступає дисперсія помилки. На відміну від Statistica, Eviews дає можливість побудови моделі стаціонарних рядів з урахуванням зміни дисперсії. Також існує можливість тестування знайденої моделі АРКС на гетероскадестичну складову за допомогою ARCH.

Щодо моделей Tarch і E-garch в Eviews версій 5.1 і вище надається можливість їх оцінки і специфікації. В Eviews також присутня можливість виправлення помилок і проведення специфікації векторної авторегресійної моделі. Під виправленням помилок розуміється дослідження коінтеграції між декількома змінними. Це можна зробити, використовуючи процедуру Йохансена. В групі тимчасових рядів ця процедура оцінює максимальну правдоподібність векторів швидкості приведення і коінтеграції.

1.1.6. Моделювання нестационарних рядів

Для звичайного моделювання нестационарних рядів застосовуються моделі АРІКС, порядок інтеграції яких є порядком взяття різниць. Застосовуючи процедуру перетворення ряду в Statistica різниці обчислюють до досягнення рядом стаціонарності. Перевірити стаціонарність можна лише за допомогою візуального аналізу, потім відбувається ідентифікація і побудова моделі АРКС. Для побудови моделі в Eviews застосовується розширений тест Дікі-Фуллера. Після отримання різниці першого та другого порядків перевірка стаціонарності виконується автоматично (необхідний порядок потрібно ввести в діалоговому вікні). Отже Eviews є більш зручним для побудови моделей АРІКС.

1.2. Типи нестационарних процесів

Стаціонарність буває слабкою і сильною. Також існують наступні назви слабкої стаціонарності:

- стаціонарність за коваріацією;
- стаціонарність другого порядку;
- стаціонарність в широкому сенсі.

Формально, стаціонарним по коваріації, називають стохастичний процес що має скінченне середнє та дисперсію, якщо для всіх k і $k-s$, ($k=0, 1, 2, \dots$; $s=0, 1, 2, \dots$) виконуються наступні умови:

- математичне сподівання не залежить від часу:

$$E[y(k)] = E[y(k-s)] = \mu = \text{const} \quad (1.1)$$

- на всьому часовому інтервалі досліджуваного процесу, дисперсія залишається постійною.

$$E\{[y(k) - \mu]^2\} = E\{[y(k - s) - \mu]^2\} = \sigma^2 = \text{const}, \quad (1.2)$$

або

$$\text{var}[y(k)] = \text{var}[y(k - s)] = \sigma^2 = \text{const} \quad (1.3)$$

— автоковаріація на всьому часовому інтервалі залишається незмінною в часі:

$$\begin{aligned} E\{[y(k) - \mu][y(k - s) - \mu]\} = \\ E\{[y(k - j) - \mu][y(k - j - s) - \mu]\} = y(k) = \text{const}, \end{aligned} \quad (1.4)$$

або

$$\begin{aligned} \text{cov}[y(k), y(k - s)] = \\ \text{cov}[y(k - j), y(k - j - s)] = y(s) = \text{const}. \end{aligned} \quad (1.5)$$

Константні значення математичного сподівання та дисперсії не є обов'язковими для сильної стаціонарності.

Виходячи з цього слабкою стаціонарністю на поведінку випадкового процесу накладаються обмеження більш строгого характеру, ніж сильною стаціонарністю. Процеси, що є стаціонарними в слабкому сенсі називаються стаціонарними по коваріації. Отже для нестаціонарності процесу досить невиконання хоча б однієї з умов (1.1)—(1.3).

Основні статистичні характеристики процесу, такі як математичне сподівання, дисперсія і коваріація, обчислюються на основі математичної моделі досліджуваного процесу (теоретичні характеристики), або конкретних значень часового ряду (вибіркові характеристики).

Для знаходження конкретних параметрів процесів та їх комбінацій, що мають вплив на стаціонарність, важливо знайти необхідні та достатні умови не стаціонарності цих процесів.

Надалі розглянемо два типи сильної стаціонарності:

- процеси зі змінним математичним сподіванням або процеси з трендом
- процеси зі змінною дисперсією або гетероскадестичні процеси

1.2.1. Інтегровані процеси

У випадку, коли $E[x(k)] \neq const$, тобто математичне сподівання змінюється в часі, то такий процес називають процесом з трендом або інтегрованим процесом (по аналогії із характером зміни сигналу на виході інтегратора) або процесом з одиничними коренями (відповідного характеристичного рівняння).

Поточне середнє значення процесу, отримане при використанні процедури цифрової фільтрації, надалі будемо називати трендом. Це може розглядатись як формула для обчислення поточного середнього значення в простих випадках.

Для виявлення наявності нелінійного детермінованого тренду можна використовувати оцінювання рівняння:

$$y(k) = a_0 + c_1 k + c_2 k^2 + \dots + c_m k^m, \quad (1.6)$$

що є поліномом порядку m відносно часу. Гіпотеза про відсутність тренду відхиляється, якщо серед коефіцієнтів $x_i, i=1 \dots m$ хоча б один є статистично значимим. Моделі випадкових трендів на основі комбінацій випадкових величин, використовують при відносно швидкій зміні напряму розвитку тренду або якщо виникають складнощі із знаходженням адекватного функціонального описання.

1.2.2. Гетероскедастичні процеси

Гетероскедастичними називають процеси із змінною в часі дисперсією, якщо ж при моделюванні та прогнозуванні у процесу на відрізку часу розглядається стала дисперсія, то процес називається гомоскедастичним.

Під час дослідження фінансово-економічних процесів, в якості міри ризику часто використовують дисперсію та стандартне відхилення, через те для данної проблеми у спеціальній літературі відводиться багато уваги. Також дисперсію та стандартне відхилення застосовують в якості міри волатильності(мінливості) процесу, при дослідженні фінансових процесів.

Дисперсія являється мірою розсіювання вимірів в технічних процесах. Оскільки вона може давати характеристику поточному стану механізмів, інтенсивностям випадкових збурень, технологічним процесам, інтенсивностям шумів і так далі, дисперсія, при дослідженні поточних та прогнозуванні майбутніх станів системи, є дуже важливим статистичним параметром.

Так як дисперсія може змінюватись у часі, математична модель повинна не тільки прогнозувати значення дисперсії на один або більше одного кроків вперед, а й коректно описувати її поведінку. Завдяки цьому покращиться якість рішень щодо управління процесами, що мають змінну в часі дисперсію. Прикладами можуть слугувати прийняття рішень при проведенні торгових операцій з акціями, або оцінка ризиків в банківській сфері. Дисперсія та стандартне відхилення в технічних системах виступають в якості мір відхилення поточних розмірів від заданих, рівня шуму і так далі. Їх використовують при аналізуванні якості продукції, оцінці точності розмірів і так далі.

Процес називається гетероскедастичним (ГСП), якщо

$$\text{var}[\varepsilon(k)] = \sigma_{\varepsilon^2} \neq \text{const} \quad (1.7)$$

Якщо, незалежно від значень факторів, залишається сталою величиною варіація кожної випадкової величини $\varepsilon(k)$ навколо її математичного сподівання, можна припускати гомоскедастичність.

Якщо ж дисперсія виявляється більш складною функцією часу, або змінюється в часі, можна стверджувати гетероскедастичність. Тобто, при створенні моделі процесу, потрібно знайти закон, за яким вона може змінюватись. Інколи застосовують припущення про відносно просто форму гетероскедастичності.

$$\sigma_{\varepsilon(k)^2} = c^2 x^2(k), \quad (1.8)$$

де c – константа, що має оцінюватись при використанні експериментальних даних та обраного методу для оцінки параметрів; $x(k)$ – випадкова змінна, що визначає дисперсію процесу. Перейдемо до простого випадку мультиплікативного збурення, отже

$$y(k+1) = \varepsilon(k+1)x(k), \quad (1.9)$$

де $y(k+1)$ – основна залежна змінна; $\varepsilon(k+1)$ – збурення, що містить скінченну дисперсію σ^2 ; $x(k)$ – незалежна змінна з відомими, на досліджуваному відрізку часу, вимірами. Послідовність $\{y(k)\}$ є білим шумом з постійною дисперсією, якщо $x(k) = x(k-1) = x(k-2) = \dots = \text{const}$. Але, якщо $\{x(k)\}$ – послідовність різних за значеннями чисел, то умовна дисперсія $y(k+1)$ при відомих значеннях $x(k)$ визначається наступним чином:

$$\text{var}[y(k+1)|x(k)] = x^2(k)\sigma^2. \quad (1.10)$$

Звідси видно, що від реалізації $x(k)$ залежить умовна дисперсія величини $y(k+1)$. Дисперсія величини $y(k)$, при відомому значенні $x(k)$, може

розглядатись як умовна, оскільки в момент k можна виміряти $x(k)$. Дисперсія $y(k+1)$ також буде великою (малою), при великій (малій) величині $x^2(k)$. Отже якщо значення послідовності $\{x(k)\}$ мають додатну кореляцію, то у послідовності $\{y(k)\}$ буде тенденція до послідовної додатної кореляції. Частіше на практиці розглядають модифіковану версію моделі типу

$$\ln(y(k)) = a_0 + a_l \ln(x(k-1)) + \ln e(k), \quad (1.11)$$

де $e(k)$ є похибкою моделі. Необхідність знаходження такої змінної $x(k)$, яка має суттєвий вплив на дисперсію процесу $y(k)$, що з практичної точки зору можливо не завжди, так як її може не існувати, є недоліком такої моделі. Також для постійності дисперсії результуючого ряду, необхідно певним чином обробляти первинний (вихідний) часовий ряд. В приведеному прикладі (1.6) припускається, що послідовність $e(k)$ з постійною дисперсією. При невиконанні цієї умови потрібно звернутись до додаткової обробки даних.

Висновки до розділу 1

Eviews та Statistica є найбільш вживаними програмами для статистичної обробки даних. Був проведений порівняльний аналіз обох пакетів, описано різницю при їх використанні і недоліки кожної з них. Statistica володіє простим графічним інтерфейсом, що значно спрощує використання, так як відсутня необхідність вивчення команд та навичок роботи з командним рядком, що є основним недоліком Eviews. До серйозних недоліків пакету відноситься відсутність тестів на нестационарність, моделювання гетероскедастичних процесів, функції моделювання процесів з трендом та моделі АРІКС є важливим недоліком Statistica.

Проаналізувавши недоліки і переваги розглянутих пакетів, сформуємо основні напрямки роботи. Необхідна система, що матиме зручний графічний інтерфейс, введення даних, а також редагування та порівняння шляхом аналізу даних. Не менш важливим є наявність тестів на стаціонарність, моделювання процесів з трендом та зручного командного інтерпретатора.

Постановка задачі

1. Виконати аналіз методів математичного моделювання нестационарних процесів ціноутворення з метою подальшого оцінювання прогнозів.
2. Дослідити статистичні тести для аналізу даних на наявність тренду і гетероскедастичності.
3. Побудувати адекватні моделі вибраних процесів ціноутворення.
4. Обчислити оцінки короткострокових прогнозів досліджуваних процесів та проаналізувати їх якість.
5. Сформулювати перспективи подальших досліджень.

РОЗДІЛ 2. МЕТОДИ МОДЕЛЮВАННЯ І ПРОГНОЗУВАННЯ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ, КРИТЕРІЇ АДЕКВАТНОСТІ

2.1. Моделі процесів з детермінованим трендом

Тренд (поточне середнє) може бути зростаючим або спадаючим, а за характером зміни в часі може бути детермінованим або стохастичним.

Детермінований тренд описують вибраною функцією, наприклад, поліномом від часу, сплайном, експонентою, комбінацією тригонометричних функцій та інше. Часто використовують поліноми від часу вигляду:

$$y(k) = a_0 + a_1 k + a_2 k^2 + \dots + a_m k^m + \varepsilon(k), \quad (2.1)$$

де k — дискретний час, який зв'язаний з неперервним реальним часом t через період реєстрації (дискретизації) даних: $t = kT$; $\varepsilon(k)$ — випадкова змінна, оцінка якої знаходиться шляхом оцінювання рівняння: $\hat{\varepsilon}(k) = e(k)$, де $e(k)$ — похибка моделі. Після проведення оцінки моделі послідовність значень $\{e(k)\}$ матиме всі коливання, що накладаються на тренд.

Також в моделюванні, для описання тренду застосовують експоненціальні та гармонічні поліноми.

Для моделювання випадкових трендів, необхідної точності опису яких не можна досягнути використовуючи детерміновані функції, використовують випадкові процеси. Але цей підхід розглядатись не буде.

Таким чином, використовуючи рівняння (1.1) для опису тренду, ми вилучаємо його з процесу, а повна модель процесу міститиме як мінімум два рівняння: рівняння (1.1) і рівняння, що описує накладені на тренд коливання (AR(p) або ARK(p,q)).

Видалити тренд з процесу можна за допомогою різниць. Перші різниці видаляють тренди першого порядку, другі різниці відповідно видаляють квадратичний тренд і т.д. Наприклад, нехай $y(k) = a_0 + a_1k$. Перші різниці цього процесу

$$\begin{aligned}\Delta y(k) &= y(k) - y(k-1) = a_0 + a_1k - [a_0 + a_1(k-1)] \\ &= a_1\end{aligned}\tag{2.2}$$

видаляють лінійний тренд. Неможливість прогнозування тренду після его видалення очевидна.

2.1.1. Тест на тренд

Щоб перевірити стаціонарність, часовий ряд спочатку потрібно візуально дослідити. Згадаємо, що властивістю слабкої стаціонарності є незмінні в часі дисперсія і коваріація ряду. Наявність лінійного чи нелінійного тренду часто можна визначити використовуючи візуальне дослідження. Стаціонарний ряд має нульовий порядок інтегрованості, записується це наступним чином: $\{y(k)\} \sim \{I(0)\}$.

Число, що вказує на необхідну кількість застосування оператора перших різниць до часового ряду з ціллю переходу до стаціонарного ряду називається порядком інтегрованості.

За визначенням часовий ряд має одиничний корінь або порядок інтеграції 1, тобто $\{y(k)\} \sim \{I(1)\}$, якщо його перші різниці $\Delta y(k) = y(k) - y(k-1)$ утворюють стаціонарний ряд $\{\Delta y(k)\} \sim \{I(0)\}$. [2]

Часовий ряд має два одиничних корені або, якщо досягнення стаціонарності вимагає обчислення його другої різниці, порядок інтеграції 2:

$$\Delta^2 y(k) = \Delta y(k) - \Delta y(k-1) == y(k) - y(k-1) - [y(k-1) - y(k-2)] == y(k) - 2y(k-1) + y(k-2), \quad (2.3)$$

де $\Delta^2 y(k) \sim I(0)$. В загальному випадку, порядок інтегрованості часового ряду може бути довільним $y(k) \sim I(int)$.

2.2. Поліноміальна регресія

Поліноміальна регресія є одним з найпоширеніших методів для опису процесів з трендом:

$$y(k) = a_0 + a_1 k + a_2 k^2 + \dots + a_m k^m + \varepsilon(k), \quad (2.4)$$

де $y(k)$ – основна (залежна) змінна процесу; $k = 0, 1, 2, \dots$ – дискретний час, який зв'язаний із реальним неперервним часом t через період дискретизації вимірів $T_s: t = k \cdot T_s$, $a_i, i = 0, \dots, m$ – коефіцієнти (параметри) моделі; m – порядок полінома, Порядок полінома визначається через кількість похідних, що обчислюються шляхом використання адекватної поліноміальної моделі процесу; Вплив на змінну випадкових збурень, похибки обчислень та надлишкових або неврахованих регресорів, характеризується випадковою змінною $\varepsilon(k)$. Досить часто модель (1) застосовують має практичне застосування завдяки можливості МНК для оцінювання параметрів та простоті визначення її структури.

2.3. Моделі випадкових трендів

Тренд розглядають як стохастичну функцію часу. Альтернативою означенню стохастичного тренду, наведеному у вступі, може бути таке: тренд будемо вважати стохастичним, якщо після достатнього згладжування випадкових коливань процесу методами цифрової та оптимальної фільтрації серед можливих моделей тренду кращим виявляється його опис у формі комбінації випадкових процесів, а не у вигляді детермінованої функції.

Одним із підходів до опису локального тренду є введення залежності коефіцієнтів моделі від часу:

$$y(k) = a(k) + d_1(k)k, \quad (2.6)$$

де $a(k)$ – локальна константа; локальний нахил тренду визначається за допомогою коефіцієнта $d_1(k)$.

Рекурсивні рівняння типу $y(k) = a_0 + y(k-1)$ або ускладнений їх варіант, являється альтернативним підходом для описування локальних трендів.

$$y(k) = a(k) + y(k-1) + \varepsilon(k). \quad (2.7)$$

Для спрощення аналізу випадкову змінну $\varepsilon(k)$ можна визначити як послідовність білого шуму, що містить відому дисперсію. $a(k)$ — змінна в часі перетину. Рівнянням, яке описує $a(k)$ необхідно доповнити Модель (3). Наприклад, $a(k) = a(k-1) + v(k)$, де $v(k)$ – випадкове збурення. Безпосередньо виміряти швидкість зростання $a(k)$ неможливо.

Однією з найпростіших моделей є модель випадкового кроку. Завдяки їй ми можемо описувати випадкові тренди у деяких випадках. Ця модель має наступний вигляд $y(k) = y(k-1) + \varepsilon(k)$ або $\Delta y(k) = \varepsilon(k)$, де $\varepsilon(k)$ є білим шумом

з нульовим середнім. Випадкова величина в даній моделі є приростом значення основної змінної. Ось як буде виглядати розв'язок цього рівняння:

$$\begin{aligned} y(1) &= y_0 + \varepsilon(1), \\ y(2) &= y(1) + \varepsilon(2) = y_0 + \varepsilon(1) + \varepsilon(2), \\ y(k) &= y_0 + \sum_{i=1}^k \varepsilon(i). \end{aligned} \quad (2.8)$$

Оскільки розв'язок (4) має незважену суму випадкових величин, його називають навласним. За допомогою конкретної реалізації випадкового процесу визначається напрям руху такого розв'язку. Далі, для отриманого розв'язку, наведемо стохастичні характеристики. Математичне сподівання $E[y(k)] = y_0 = \text{const}$. Умовне математичне сподівання для $y(k+1) = y(k) + \varepsilon(k+1)$, виглядає таким чином:

$$\begin{aligned} E_k[y(k+1)|y(k), y(k-1), \dots, \varepsilon(k), \varepsilon(k-1), \dots] &= E[y(k) + \\ &\varepsilon(k+1) = y(k)]. \end{aligned} \quad (2.9)$$

На основі розв'язку наведемо функцію прогнозування на довільну кількість кроків:

$$\begin{aligned} y(k) &= y_0 + \sum_{i=1}^k \varepsilon(i), \\ y(k+s) &= y_0 + \sum_{i=1}^{k+s} \varepsilon(i) = y_0 + \sum_{i=1}^k \varepsilon(i) + \sum_{i=k+1}^{k+s} \varepsilon(i) = \\ &= y(k) + \sum_{i=k+1}^{k+s} \varepsilon(i), \\ E_k[y(k+s)] &= y(k) + E_k[\sum_{i=k+1}^{k+s} \varepsilon(i)] = y(k). \end{aligned} \quad (2.10)$$

Тобто значення оцінки прогнозу не залежить від кількості кроків прогнозування. Наступною формулою, для розв'язку $y(k)$, знаходимо значення дисперсії:

$$\begin{aligned} \text{var}[y(k)] &= E\{[y(k) - \mu_y][y(k) - \mu_y]\} = E\{[y(0) + \\ &\sum_{i=1}^k \varepsilon(i) - y(0)][y(0) + \sum_{i=1}^k \varepsilon(i) - y(0)]\} = E\{[\varepsilon(1) + \\ &\varepsilon(2) + \dots + \varepsilon(k)][\varepsilon(1) + \varepsilon(2) + \dots + \varepsilon(k)]\} = k\sigma_\varepsilon^2. \end{aligned} \quad (2.11)$$

Для моменту часу $(k - s)$:

$$\begin{aligned} \text{var}[y(k - s)] &= E\{[\varepsilon(1) + \varepsilon(2) + \dots + \varepsilon(k - s)][\varepsilon(1) + \dots + \\ &\varepsilon(k - s)]\} = (k - s)\sigma_\varepsilon^2. \end{aligned} \quad (2.12)$$

Отже, $\text{var}[y(k)] \neq \text{const}$ і $\lim_{k \rightarrow \infty} \text{var}[y(k)] = \infty$.

Автокореляційна функція цього процесу буде виглядати наступним чином:

$$\rho(s) = \frac{y(s)}{\sqrt{\text{var}[y(k)]\text{var}[y(k-s)]}} = \frac{(k-s)\sigma^2}{\sqrt{k\sigma^2(k-s)\sigma^2}} = \sqrt{\frac{k-s}{k}}. \quad (2.13)$$

Тобто, для процесу випадкового кроку, АКФ є повільно спадаючою функцією.

Модель випадкового кроку плюс дрейф (зміщення або перетин). У даному випадку додамо константу a_0 до моделі випадкового кроку:

$$y(k) = a_0 + y(k - 1) + \varepsilon(k). \quad (2.14)$$

При відомій початковій умові $y(0) = y_0$ розв'язок рівняння (2.3.9) має вигляд:

$$y(k) = y_0 + a_0 k + \sum_{i=1}^k \varepsilon(k). \quad (2.15)$$

Отже, лінійний детермінований тренд $a_0 k$ і стохастичний (нерегулярний) тренд $\sum \varepsilon(k)$. є двома нестационарними компонентами, що впливають на $y(k)$. Математичне сподівання і умовне математичне сподівання мають вигляд:

$$E[y(k)] = y_0 + a_0 k, \quad E_k[y(k+s)] = y_0 + a_0(k+s). \quad (2.16)$$

З метою отримання функції прогнозування приведемо рівняння для моменту $(k+s)$:

$$y(k+s) = y_0 + a_0(k+s) + \sum_{i=1}^{k+s} \varepsilon(k) = y(k) + a_0 s + \sum_{i=1}^k \varepsilon(k+i). \quad (2.17)$$

і умовне математичне сподівання

$$E_k[y(k+s)|k] = y(k) + a_0 s. \quad (2.18)$$

Таким чином, від функції прогнозування для випадкового кроку (4), отримана функція відрізняється наявністю складової $a_0 s$.

Модель випадкового кроку з додатковою шумовою складовою. Залежна змінна $y(k)$ в цій моделі визначається сумою стохастичного тренду та випадкової компоненти, тобто

$$\begin{aligned} y(k) &= \mu(k) + \eta(k), \\ \mu(k) &= \mu(k-1) + \varepsilon(k), \end{aligned} \quad (2.19)$$

де $\{\eta(k)\}, \{\varepsilon(k)\}$ – незалежні процеси білого шуму з дисперсіями σ_η^2 та σ_ε^2 , відповідно; $E[\varepsilon(k)\eta(k-s)] = 0, \forall k, s$. Для початкової умови $\mu(0) = \mu_0$ розв’язок рівняння виглядає наступним чином: $\mu(k) = \mu_0 + \sum_{i=1}^k \varepsilon(i)$, що є випадковим трендом для $y(k)$. Тепер $y(k) = \mu_0 + \sum_{i=1}^k \varepsilon(i) + \eta(k)$. Застосовуючи початкову умову $y(0) = y_0 = \mu_0 + \eta_0$, отримаємо розв’язок:

$$y(k) = y_0 - \eta_0 + \sum_{i=1}^k \varepsilon(i) + \eta(k). \quad (2.20)$$

Безумовне середнє процесу $\{y(k)\}$ є константою $E[y(k)] = E[y(k+s)] = y_0 - \eta_0$. Істо шумова компонента $\{\eta(k)\}$ впливає на процес в кожний момент часу. Дисперсія процесу $\{y(k)\}$ залежить від часу, оскільки $Var[y(k)] = k\sigma_\varepsilon^2 + \sigma_\eta^2$ і $Var[y(k-s)] = (k-s)\sigma_\varepsilon^2 + \sigma_\eta^2$. Теж саме ми можемо спостерігати і у інших моделей із стохастичним трендом $Var[y(k)] \rightarrow \infty$ при $k \rightarrow \infty$. Порівняно з моделью випадкового кроку, у данному випадку коефіцієнт кореляції між $y(k)$ та $y(k-s)$ менший. Про це свідчить наявність шумової компоненти. Коваріація між $y(k)$ і $y(k-s)$ визначається наступним чином:

$$\begin{aligned} cov[y(k)y(k-s)] &= E\{[y(k) - y_0 - \eta_0][y(k-s) - y_0 - \eta_0]\} \\ &= E\{[\varepsilon(1) + \varepsilon(2) + \dots + \varepsilon(k) + \eta(k)][\varepsilon(1) + \varepsilon(2) + \dots \\ &\quad + \varepsilon(k-s) + \eta(k-s)]\} = (k-s)\sigma_\varepsilon^2. \end{aligned} \quad (2.21)$$

Так як $\{\varepsilon(k)\}$ і $\{\eta(k)\}$ – незалежні послідовності. Отже, автокореляційну функцію $\rho(s)$ можна визначити наступним виразом:

$$p(s) = \frac{(k-s)\sigma_\varepsilon^2}{\sqrt{k\sigma_\varepsilon^2 + \sigma_\eta^2[k-s]\sigma_\varepsilon^2 + \sigma_\eta^2}}. \quad (2.22)$$

Проаналізувавши два вирази значимо, що для процесу випадкового кроку з додатковим шумом коефіцієнти автокореляції при $\sigma_\eta^2 > 0$ є завжди меншими.

Скориставшись рівнянням (2.20) запишемо функцію прогнозування на s кроків:

$$\begin{aligned} y(k+s) &= y_0 - \eta_0 + \sum_{i=1}^{k+s} \varepsilon(i) + \eta(k+s) = \\ &= y(k) - \eta(k) + \sum_{i=1}^s \varepsilon(k+i) + \eta(k+s). \end{aligned}$$

Умовне математичне сподівання для $y(k+s)$: $E_k[y(k+s)] = y(k) - \eta(k)$.

Отже робимо висновок про присутність тренду та нерегулярної компоненти у моделі випадкового кроку з шумом, а поточне значення $y(k)$, з якого складається прогноз, зменшується на випадкову величину $\eta(k)$. Постійною складовою $\{y(k)\}$ є стохастичний тренд у вигляді $\sum \varepsilon(k)$.

Модель випадкового кроку з шумом та дрейфом виглядає таким чином:

$$\begin{aligned} y(k) &= \mu(k) + \eta(k), \\ \mu(k) &= \mu(k-1) + \varepsilon(k), \end{aligned} \quad (2.23)$$

де a_0 – константа; $\{\varepsilon(k)\}, \{\eta(k)\}$ – незалежні процеси білого шуму. У розглянутій моделі тренд $\mu(k)$ має детерміновану a_0 та випадкову складову $\varepsilon(k)$.

Запишемо розв'язок для $\mu(k)$: $\mu(k) = \mu_0 + a_0 k + \sum_{i=1}^k \varepsilon(i)$, які дають нам перехід до

наступного виразу для $y(k)$: $y(k) = \mu_0 + a_0 k + \sum_{i=1}^k \varepsilon(i) + \eta(k)$.

Запишемо рівняння для $y(k)$, врахувавши початкову умову $y_0 = \mu_0 + \eta_0$:

$$y(k) = y_0 - \eta_0 + a_0 k + \sum_{i=1}^k \varepsilon(i) + \eta(k). \quad (2.24)$$

Отже, $y(k)$ є сумою детермінованого і стохастичного трендів і чисто випадкової складової $\eta(k)$. Білий шум не обов'язково повинен висутпати в ролі шумової складової. Рівняння (2.24) можна переписати у вигляді:

$$y(k) = \mu_0 + a_0 k + \sum_{i=1}^k \varepsilon(i) + A(L) + \eta(k), \quad (2.25)$$

де $A(L)$ – поліном відносно оператора затримки L ; $A(L)\eta(k)$ – стаціонарний шумовий процес. Рівняння (2.25) називають моделлю узагальненого тренду з нерегулярною складовою. [4]

Далі запишемо модель локального лінійного тренду, який містить три процеси випадкового кроку з шумом:

$$\begin{aligned} y(k) &= \mu(k) + \eta(k), \\ \mu(k) &= \mu(k-1) + \lambda(k) + \varepsilon(k), \\ \lambda(k) &= \lambda(k-1) + v(k), \end{aligned} \quad (2.26)$$

де $\{\eta(k)\}$, $\{\varepsilon(k)\}$, $\{v(k)\}$ – незалежні процеси білого шуму. Шумова складова та процес випадкового кроку в розглянутому випадку призводять до змін тренду, тобто складовими $\Delta\mu(k)$ є процес випадкового кроку $\lambda(k)$ і білий шум $\varepsilon(k)$. Те, що розглянуті вище випадки є окремими випадками ЛЛТ, можна легко показати.

Перейдемо до знаходження розв'язку $y(k)$. Але зараз приведемо розв'язок для $\lambda(k)$:

$$\lambda(k) = \lambda_0 + \sum_{i=1}^k v(i), \quad (2.27)$$

а також для $\mu(k)$:

$$\mu(k) = \mu(k-1) + \lambda_0 + \sum_{i=1}^k v(i) + \varepsilon(k), \quad (2.28)$$

або

$$\mu(k) = \mu_0 + \sum_{i=1}^k \varepsilon(i) + k(\lambda_0 + v(1)) + (k-1)v(2) + (k-2)v(3) + \dots + v(k).$$

Оскільки $y_0 = \mu_0 + \eta_0$, то розв'язок для $y(k)$ має вигляд:

$$y(k) = y_0 + [\eta(k) - \eta_0] + \sum_{i=1}^k \varepsilon(i) + k(\lambda_0 + v(1)) + (k-1)v(2) + \dots + v(k). \quad (2.3.29)$$

У цьому рівнянні можна спостерігати властивості всіх інших моделей. Кожний елемент послідовності $\{y(k)\}$ містить детермінований тренд $(k(\lambda_0 + v(1)) + (k-1)v(2) + \dots + v(k))$, стохастичний тренд $(\sum \varepsilon(i))$ та нерегулярну компоненту $\eta(k)$. Якщо при роботі з моделю ЛЛТ, розглядати її узагальнену форму, то нерегулярна компонента буде визначена за допомогою члена $A(L)\eta(k)$. У розглянутому випадку простежується залежність детермінованого тренду від поточних та минулих значень послідовності $\{v(k)\}$. Додатнім коефіцієнт при k , буде тільки якщо сума $(\lambda_0 + v(1) + \dots + v(k))$ буде додатною в момент k . В узагальненому випадку для деяких k ця сума може бути додатною, а для деяких від'ємною, що свідчить про можливість наявності як додатного так і від'ємного нахилу відрізків тренду, що для багатьох фінансових процесів є характерним. Функція прогнозування на s кроків для ЛЛТ:

$$y(k+s) = y_0 + [\eta(k) - \eta_0] + \sum_{i=1}^k \varepsilon(i) + k(\lambda_0 + v(1)) + \\ + (k+s-1)v(2) + (k+s-2)v(3) + \dots + v(k+s) \quad (2.29)$$

або

$$y(k+s) = y(k) + [\eta(k+s) - \eta(k)] + \sum_{i=1}^k \varepsilon(i) + s[\lambda_0 + v(1) + v(2) + \dots + v(k)] + \\ + \sum_{i=1}^s (s+1-i)v(k+i). \quad (2.30)$$

Умовне математичне сподівання має вигляд:

$$E_k [y(k+s)] = [y(k) - \eta(k)] + s[\lambda_0 + v(1) + v(2) + \dots + v(k)] \quad (2.31)$$

Нахил тренду визначається сумою $s[\lambda_0 + v(1) + v(2) + v(3) + \dots + v(k)]$.

Для виконання обчислювальних експериментів при побудові моделі стохастичних трендів на основі фактичних даних, необхідно використання методу Монте-Карло. У випадкових процесах метод Монте-Карло застосовується для генерації множини їх допустимих реалізацій, при наявності параметрів, оцінених апіорно. Ці оцінки можна літримати використовуючи експертну оцінку або досліджуючи фактичні дані. Побудовані на основі таких методів алгоритми імітаційного моделювання дозволяють нам отримувати прийнятні оцінки оцінки середньострокових прогнозів при дослідженні трендів, що мають складну форму.

2.4. Моделі гетероскедастичних процесів

Прогнозування на основі умовного та безумовного середнього .

Нехай проводиться оцінка моделі стаціонарного процесу авторегресії із ковзним середнім (АРКС) типу (АР(1):

$$\begin{aligned} y(k) &= a_0 + a_1 y(k-1) + \varepsilon(k), |a_1| < 1, \\ E[\varepsilon(k)] &= 0, \quad E[\varepsilon^2(k)] = \sigma_\varepsilon^2, \quad E[\varepsilon(k)\varepsilon(k-l)] = 0, l > 0, \\ E[\varepsilon(k)y(k)] &\approx \sigma_\varepsilon^2. \end{aligned} \quad (2.32)$$

Потрібно здійснити оцінку прогнозу величини $y(k+1)$. Умовним математичним сподіванням для процесу АР(1) є

$$E_k[y(k+1)] = a_0 + a_1 y(k) \quad (2.33)$$

При застосуванні цього умовного середнього для прогнозування значення $y(k+1)$, дисперсію похибки прогнозу можна визначити як:

$$E_k\{[y(k+1) - a_0 - a_1 y(k)]^2\} = E_k[\varepsilon^2(k+1)] = \sigma_\varepsilon^2. \quad (2.34)$$

Знайдемо безумовне середнє, тобто,

$$E[y(k)] = a_0 + a_1 E[y(k-1)] + E[\varepsilon(k)], \quad (2.35)$$

і при $E[y(k)] = E[y(k-1)]$ отримуємо, що $E[y(k)] = \frac{a_0}{1-a_1}$. На нескінченному часовому інтервалі, для послідовності $\{y(k)\}$, воно відображає середнє значення,

яке, в даному випадку, рівняється $a_0/(1-a_1)$, до цього ми приходимо завдяки розв'язку AP(1), що має наступний вигляд:

$$y(k) = a_0 \sum_{i=0}^{\infty} a_1^i + y_0 a_1^k + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i),$$

а при $k \rightarrow \infty$ маємо:

$$y(k) = a_0 \sum_{i=0}^{\infty} a_1^i + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i) = \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i), |a_1| < 1. \quad (2.36)$$

Наступним виразом визначимо дисперсію похибки безумовного прогнозу:

$$\begin{aligned} E\{[y(k+1) - a_0/(1-a_1)]^2\} &= \\ &= E\{[\varepsilon(k+1) + a_1 \varepsilon(k) + a_1^2 \varepsilon(k-1) + a_1^3 \varepsilon(k-2) + \dots]^2\} = \\ &= \sigma^2/(1-a_1^2), \quad |a_1| < 1. \end{aligned} \quad (2.37)$$

При знаходженні безумовного середнього, у результаті використання рівняння першого порядку в правій частині отримуємо послідовність $\{\varepsilon(k)\}$. Робимо висновок про більшу дисперсію безумовного прогнозу в порівнянні з умовним, так як $1/(1-a_1^2) > 1$, в зв'язку з цим віддають перевагу умовному прогнозуванню. Тенденцію зміни дисперсії послідовності $\{\varepsilon(k)\}$ в часі можна описати із застосуванням моделі АРКС. Візьмемо модель першого порядку $y(k) = a_0 + a_1 y(k-1) + \varepsilon(k)$. $\{\hat{\varepsilon}(k)\}$, у моделі першого порядку, є оцінкою похибок моделі. Для основної змінної, умовну дисперсію отримаємо наступним чином:

$$\text{Var}[y(k+1)/y(k)] = E_k\{[y(k+1) - a_0 - a_1 y(k)]^2\} = E_k[\hat{\varepsilon}^2(k+1)]. \quad (2.38)$$

Покладемо тепер, що умовна дисперсія – величина змінна. Одним із простих підходів до описання такої змінної величини є застосування моделі типу $AR(q)$ до квадратів оцінок залишків. Наприклад,

$$\hat{\varepsilon}^2(k) = \alpha_0 + \alpha_1 \hat{\varepsilon}^2(k-1) + \alpha_2 \hat{\varepsilon}^2(k-2) + \dots + \alpha_q \hat{\varepsilon}^2(k-q) + v(k), \quad (2.39)$$

де $v(k)$ – процес білого шуму з нульовим середнім для адекватної моделі.

При нульових значеннях коефіцієнтів $\alpha_1, \alpha_2, \dots, \alpha_q$, отмуємо константну оцінку дисперсій. В іншому випадку, для опису умовної дисперсії $y(k)$ використовуємо рівняння (2.39). Застосуємо його для прогнозування умовної дисперсії на один крок наступним чином:

$$E_k[\hat{\varepsilon}^2(k+1)] = \alpha_0 + \alpha_1 \hat{\varepsilon}^2(k-1) + \alpha_2 \hat{\varepsilon}^2(k-2) + \dots + \alpha_q \hat{\varepsilon}^2(k-q).$$

Оскільки (2.39) може бути побудоване при умові, що $Var[y(k)] \neq const$, то (2.39) називають авторегресійним умовно гетероскедастичним (АРУГ) рівнянням [4]. Залишки (збурення) $\varepsilon(k)$, із рівняння (2.39), можна знайти при застосуванні наступних рівнянь: рівнянь регресії, авторегресії або авторегресії із ковзним середнім низького порядку.

Порівняно з рівнянням (2.39) існують і більш складні описання поведінки дисперсії. Наприклад, майже ніколи наперед невідомо як впливає збурення на процес — адитивно чи мультиплікативно. Оскільки ми не можемо спрогнозувати, адитивним чи мультиплікативним буде вплив збурення на процес, його можна подати у вигляді моделі, що має мультиплікативну форму:

$$\varepsilon^2(k) = v^2(k) [\alpha_0 + \alpha_1 \varepsilon^2(k-1)], \quad (2.40)$$

де $v(k)$ – мультиплікативне збурення у формі білого шуму причому $\{v(k)\} \sim (0,1)$, з нульовим середнім та одиничною дисперсією, так як $\{v(k)\} \sim (0,1)$, змінні $\varepsilon(k-1)$ і $v(k)$ — статистично незалежні величини. Для того, щоб показати некорельованість елементів послідовності $\{\varepsilon(k)\}$ та наявність у них нульового середнього, для $\varepsilon(k)$ потрібно знайти умовне математичне сподівання. Оскільки $E[v(k)] = 0$, то

$$\begin{aligned} E[\varepsilon(k)] &= E[v(k)(\alpha_0 + \alpha_I \varepsilon^2(k-1))^{1/2}] = \\ &= E[v(k)] E[\alpha_0 + \alpha_I \varepsilon^2(k-1)]^{1/2} = 0. \end{aligned} \quad (2.41)$$

При $E[v(k)v(k-1)] = 0$ маємо, що

$$E[\varepsilon(k)\varepsilon(k-i)] = 0, i \neq 0. \quad (2.42)$$

Також знайдемо безумовну дисперсію:

$$E[\varepsilon^2(k)] = E[v^2(k)(\alpha_0 + \alpha_I \varepsilon^2(k-1))] = E[v^2(k)] E[\alpha_0 + \alpha_I \varepsilon^2(k-1)]. \quad (2.43)$$

Так як $\sigma_{v,2} = 1$, а безумовна дисперсія змінної $\varepsilon(k)$ така ж як і для $\varepsilon(k-1)$, тобто $E[\varepsilon^2(k)] = E[\varepsilon^2(k-1)]$, то її значення приймаються:

$$\sigma_{v,2} = 1, \quad (2.44)$$

Таким чином, показаний у рівнянні (2.40) процес, не впливає на дисперсію та безумовне середнє. Аналогічними діями доводиться, що умовне середнє процесу $\varepsilon(k)$ рівнятиметься нулю. При умові, що $v(k)$ і $\varepsilon(k-1)$ — незалежні

величини , а також $E[v(k)] = 0$, умовне середнє для $\varepsilon(k)$ знаходиться наступним чином:

$$E_{k-1}[\varepsilon(k) | \varepsilon(k-1), \varepsilon(k-2), \dots] = E_{k-1}[v(k)] E_{k-1}[\alpha_0 + \alpha_1 \varepsilon^2(k-1)]^{1/2} = 0.$$

Очевидно, що описаний виразом (2.40) процес, впливає на умовну дисперсію. Так як $\sigma_{v^2} = 1$, умовна дисперсія для $\varepsilon(k)$ виражається наступним чином:

$$E_{k-1}[\varepsilon^2(k) | \varepsilon(k-1), \varepsilon(k-2), \dots] = \alpha_0 + \alpha_1 \varepsilon^2(k-1). \quad (2.45)$$

Простежуємо залежність дисперсії змінної $\varepsilon(k)$ від значення $\varepsilon^2(k-1)$. При прийнятті великого значення останньою змінною, велике значення матиме і умовна дисперсія в момент k . Рівняння (2.45) хоч і являється авторегресією першого порядку, але на його коефіцієнти α_0, α_1 повинні накладатись певні обмеження. Щоб мати змогу стверджувати додатність дисперсії, α_0, α_1 коефіцієнти обидва мають бути додатними. Також, для стійкості процесу авторегресії необхідним є виконання умови $0 < \alpha_1 < 1$.

Рівняння (2.42)-(2.45) ілюструють загальні властивості будь-якого процесу АРУГ. Структура АРУГ має наступні особливості: умовне та безумовне середнє значення похибок цієї моделі дорівнює нулю, послідовність $\varepsilon(k)$ є послідовно некорельованою, так як $E[\varepsilon(\varepsilon(k-s))] = 0$ для $\forall s \neq 0$. В данном випадку слід виділити зв'язок елементів послідовності $\{\varepsilon(k)\}$ між собою, через умовну дисперсію (їх другий центральний момент). Дисперсія змінної $\varepsilon(k)$ має тенденцію до зростання, якщо має суттєве відхилення від нуля значення елементу $\varepsilon(k-1)$. Отже умовна гетероскедастичність $\{\varepsilon(k)\}$ каже нам про відношення до процесу типу АРУГ послідовності $\{y(k)\}$.

Далі розглянемо засіб впливу на послідовність $\{y(k)\}$ структури похибки. Про взаємодію автокореляційних параметрів послідовності $\{y(k)\}$ свідчить велика кількість обчислювальних експериментів. Пояснюється це наступним чином, при зростанні модуля $v(k)$, збільшується дисперсія $\{\varepsilon(k)\}$; при збільшенні коефіцієнта α_1 , збільшується вплив на дисперсію. Також при збільшенні параметру авто регресії a_1 , збільшується вплив на $y(k)$ правої частини. Дисперсія середнього значення послідовності буде тим більша, чим більша буде тенденція до зміни середнього значення послідовності $\{y(k)\}$.

Для послідовності $\{y(k)\}$ запишемо умовні значення середнього та дисперсії:

$$E_{k-1}[\varepsilon^2(k) | \varepsilon(k-1), \varepsilon(k-2), \dots] = \alpha_0 + \alpha_1 \varepsilon^2(k-1).$$

і

$$\begin{aligned} \text{Var}[y(k) | y(k-1), y(k-2), \dots] &= E_{k-1}[y(k) - a_0 - a_1 y(k-1)]^2 = \\ &= E_{k-1}[\varepsilon^2(k)] = \alpha_0 + \alpha_1 \varepsilon^2(k-1). \end{aligned} \quad (2.45)$$

Так як α_1 і $\varepsilon^2(k-1)$ не приймають від'ємних значень, α_0 становиться мінімальним значенням умовної дисперсії. Будь-яке ненульове значення змінної $\varepsilon(k-1)$ позитивно зв'язане з умовною дисперсією змінної $y(k)$ через коефіцієнт α_1 . Коефіцієнт a_1 позитивно зв'язує ненульове значення змінної $\varepsilon(k-1)$ з умовною дисперсією змінної $y(k)$. Розглянемо різницеве рівняння, що описує процес $\{y(k)\}$. Завдяки розв'язку такого рівняння можна отримати безумовні значення середнього та дисперсії для $y(k)$. Так буде виглядати розв'язок для рівняння першого порядку:

$$y(k) = \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i). \quad (2.26)$$

Оскільки $E[\varepsilon(k) = 0, \forall k]$, то для останнього рівняння безумовне математичне сподівання буде виглядати наступним чином:

$$E[y(k)] = \frac{a_0}{1 - a_1}. \quad (2.47)$$

Використовуючи рівняння (2.46) можна отримати безумовну дисперсію для $y(k)$:

$$\text{Var} [y(k)] = \sum_{i=0}^{\infty} a_1^{2i} \text{Var} [\varepsilon(k-i)]. \quad (2.48)$$

Безумовна дисперсія для $\varepsilon(k)$ є величиною постійною, тобто $\text{Var} [\varepsilon(k)] = \text{Var} [\varepsilon(k-1)] = \text{Var} [\varepsilon(k-2)] = \dots = \alpha_0 / (1 - \alpha_1)$, враховуючи це:

$$\text{Var} [y(k)] = \frac{\alpha_0}{(1 - \alpha_1)} \frac{1}{(1 - a_1^2)}. \quad (2.49)$$

Після розширення модель гетероскедастичного процесу приймає наступний вигляд:

$$\varepsilon(k) = v(k) \left(\alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon^2(k-i) \right)^{1/2}.$$

В цьому рівнянні на змінну $\varepsilon(k)$ впливають всі значення від $\varepsilon(k-1)$ до $\varepsilon(k-q)$, а тому умовну дисперсію можна розглядати як процес авторегресії порядку q .

2.5. Моделі стаціонарних процесів

На цьому етапі перед нами постає необхідність вибору структури моделей кандидатів. Надалі опишемо компоненти, що включає в себе структура моделі. Порядком моделі є найвищий порядок рівнянь, що його утворюють. Лагом називають час запізнення по входу, його оцінка також є одним із компонентів. Наступними компонентами є можливі нелінійності та зовнішні збурення і їх типи. Типи зовнішніх збурень бувають мультиплікативними, неперервними, випадковими чи детермінованими, адитивними .

2.5.1 Модель авторегресії

«Пам'ять» процесу пояснює існування авто регресії у природі. Проявляється вона в визначенні поточного стану попередніми станами . Наприклад, поточний стан будь-якого технологічного процесу залежить від стану цього процесу в попередні моменти часу.

Вище описане можна виразити наступною формулою:

$$y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + \dots + a_n y(k-n) \quad (2.50)$$

Далі потрібно дослідити та обчислити автокореляційну функцію змінної $y(k)$ для того, щоб оцінити необхідність вводити регресійну складову в рівняння регресії.

Використовуючи автокореляційну функцію можливо визначити порядок авторегресії. Порядок авторегресії визначається числом коефіцієнтів, що є відмінними від нуля.

Далі потрібно провести попередню оцінку порядку авто регресійної частини моделі. Під оцінкою розуміється визначення кількості затриманих в часі значень, необхідних для описання процесу. Для цього використовують АКФ та ЧАКФ. При цьому оцінка порядку АКФ є менш чіткою за оцінку порядку ЧАКФ

Наступним виразом обчислюється вибіркова АКФ:

$$r_y(s) = r_{y(k)y(k-s)} = \frac{1}{N-1} \frac{\sum_{k=s+1}^N \{ [y(k) - \bar{y}][y(k-s) - \bar{y}] \}}{\sigma_y^2}, \quad (2.51)$$

де σ_y^2 – вибіркова дисперсія змінної $y(k)$; \bar{y} – середнє значення вибірки даних.

Для визначення порядку аторегресійної частини моделі необхідно порахувати число відмінних, в статистичному сенсі, від нуля коефіцієнтів АКФ. Стаціонарним процесом називається процес, що має постійні значення дисперсії, коваріації та середнього значення. Коефіцієнти $r_y(s)$ мають нормальний розподіл та нульове середнє, якщо процес є стаціонарним.

Коефіцієнти ЧАКФ володіють чіткішим порівняно з АКФ відображенням зв'язку між окремими значеннями основної змінної. Це пояснюється тим, що часткова АКФ включає вплив величин $y(k-1) \dots y(k-s+1)$ між значеннями $y(k)$ $y(k-s)$. Під вибілковими коефіцієнтами ЧАКФ розуміються коефіцієнти, що були знайдені з використанням вибірки даних. Ці коефіцієнти обчислюються простим методом. Опишемо суть цього методу:

1) Використовуючи відхилення основної змінної формуємо додатковий часовий ряд:

$$\{y'(k)\} = \{y(k)\} - \mu \quad (2.52)$$

де μ – середнє значення ряду.

2) Формується рівняння першого порядку

$$y'(k) = \Phi_{11} y'(k-1) + e(k), \quad (2.53)$$

де під $e(k)$ розуміється похибка моделі. В такому рівнянні, між $y(k)$ та $y(k-1)$, Φ_{11} виступає в ролі коефіцієнта АКФ та ЧАКФ. Формуємо рівняння другого порядку з метою оцінки двох коефіцієнтів:

$$y'(k) = \Phi_{11} y'(k-1) + \Phi_{22} y'(k-2) + e(k), \quad (2.54)$$

де Φ_{22} – виступає в ролі коефіцієнта ЧАКФ між $y(k)$ та $y(k-2)$.

Також є можливим обчислення коефіцієнтів ЧАКФ застосовуючи коефіцієнти АКФ з використанням наступних виразів:

$$\Phi_{11} = r(1), \quad \Phi_{22} = \frac{r_2 - r_1^2}{1 - r_1^2}, \quad \Phi_{ss} = \frac{r_s - \sum_{j=1}^{s-1} \Phi_{s-1,j} r_{s-j}}{1 - \sum_{j=1}^{s-1} \Phi_{s-1,j} r_j} \quad (2.55)$$

2.5.2 Модель ковзного середнього

Під ковзним середнім розуміється інструмент згладжування цінових рядів. Найчастіше його використовують для відображення цін на різні види сировини, змін котирувань акцій на біржі і так далі. Серед ковзних середніх, найпоширенішими на практиці є експотентні, зважені та прості.

Наведемо приклад, звернувшись до цін на акції. Проста ковзна середня може визначатись як ціна відкриття акцій за певну фіксовану кількість днів. 20-денна відповідно вказуватиме на значення середньої ціни за останні 20 відкриттів. Наведемо формулу графіка просто ковзного середнього для визначення кожної точки лінії графіка.

$$MA(k) = \frac{y(k-m) + y(k-(m-1)) + \dots + y(k)}{m} \quad (2.56)$$

Набір усереднюючих значень постійно рухається у часі, саме це і означає термін ковзного середнього. Допустимо нам потрібно відобразити тенденції зміни цін і провести згладжування їх несуттєвих коливань. Саме для цього ми і застосуємо ковзне середнє. Проте зазвичай зміни ковзної середньої відбуваються в певному горизонтальному діапазоні, при відсутності яскраво вираженої тенденції.

Досягти згладжування властивостей ковзної середньої є можливим при використанні лага в інформації. Буде спостерігатись відставання графіків ковзних середніх від відповідних їм змін у цінових рядах. Це пояснюється тим, що ковзна середня дорівнює середньому значенню минулих цін. Це породжує велику кількість помилкових сигналів на ринках, де виражений тренд відсутній.

Наведемо загальну формулу для зваженого ковзного середнього, яка буде виглядати наступним чином:

$$MA(k) = \frac{\sum_{i=1}^N w_i \cdot y(k-i+1)}{\sum_{i=1}^N w_i} \quad (2.57)$$

де N – розмір вікна ковзного середнього, w_i – вагові коефіцієнти, y – часовий ряд вхідних даних.

Загальна формула для простого (арифметичного) ковзного середнього буде виглядати наступним чином:

$$MA(k) = \frac{\sum_{i=1}^N y(k-i+1)}{N} \quad (2.58)$$

Тобто всі вагові коефіцієнти, у випадку простого ковзного середнього, будуть мати однакову(одиничну) вагу.

У випадку експотенційного ковзного середнього вагові коефіцієнти розподілені за експотенційним законом. При виконанні аналізу даних із застосуванням ЕКС спостерігається експотенційне зменшення ваг з віддаленістю від поточного періоду. Ситуація може бути і оберненою, якщо розглядати загальні випадки, але їх ніхто не розглядає.

$$EMA(k) = \frac{\sum_{i=1}^N w_i \cdot y(k-i+1)}{\sum_{i=1}^N w_i} = \frac{w_1 \cdot y(k) + w_2 \cdot y(k-1) + \dots + w_n \cdot y(k-n)}{\sum_{i=1}^N w_i} \quad (2.59)$$

Вибір вагових коефіцієнтів експоненційного ковзного середнього.

Підхід № 1.

Вагові коефіцієнти експоненційного ковзного середнього обчислюються за формулами:

$$\begin{aligned} w_1 &= (1-\alpha), \\ w_1 &= (1-\alpha), \\ w_n &= (1-\alpha)^n. \end{aligned} \quad (2.60)$$

При чому $\alpha = \frac{2}{n+1}$, де n – розмір вікна.

В розглянутому випадку відбувається збільшення випадкових ваг ЕКС порівняно з поточним періодом, як показано на рис. 2.1.

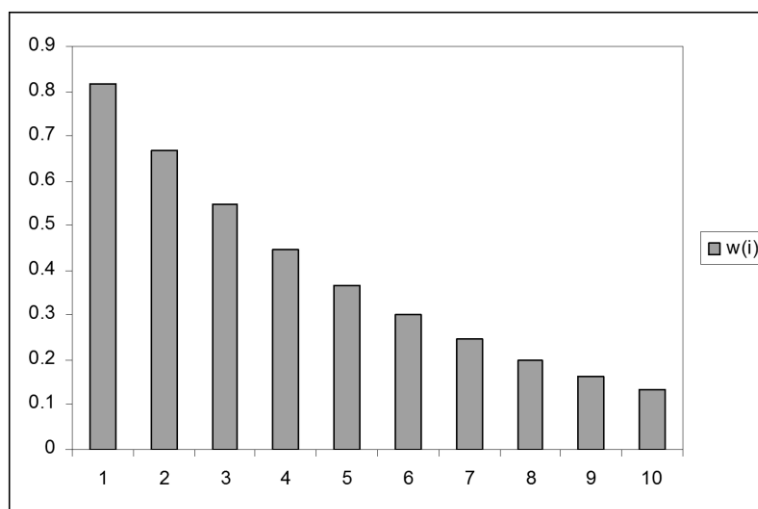


Рисунок 2.1 — Випадок коли ваги ЕКС збільшуються з віддаленістю від поточного періоду

Підхід №2.

Але якщо розглядати практичні ситуації очевидно стає велика цінність свіжої інформації про досліджуваний процес. Виходячи з цього для обчислення ваг ЕКС потрібно використовувати наступну формулу:

$$w_i = (1 - \alpha)^{n-i+1} \quad (2.61)$$

тобто $w_1 = (1 - \alpha)^n$; $w_2 = (1 - \alpha)^{n-1}$; ...; $w_n = (1 - \alpha)^2$; $w_n = (1 - \alpha)$, де $\alpha = \frac{2}{n+1}$, а n – розмір вікна ковзного середнього. На рис. 2.2 зображено гістограму вагових коефіцієнтів.

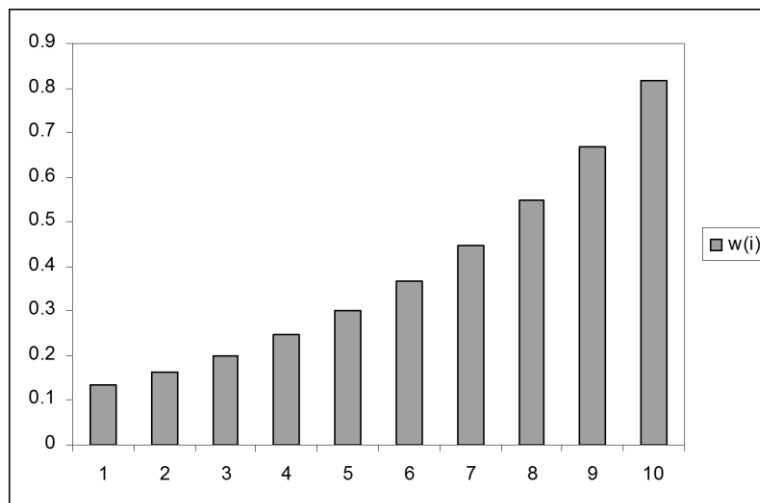


Рисунок. 2.2 — Випадок коли ваги ЕКС зменшуються з віддаленістю від поточного періоду

Алгоритм побудови $APKC(p,q)$ у випадку побудови КС на основі залишок $AP(p)$ рівняння моделі.

1. Визначення P – порядку авто регресійної складової.
2. Обчисліть АКФ та ЧАКФ для
3. По отриманим значенням АКФ та ЧАКФ визначте P – порядок авто регресійної складової.
4. Визначення q – порядку КС.
5. Обчисліть коефіцієнти моделі $AP(p)$: .
6. Обчисліть коефіцієнти АКФ та ЧАКФ залишків моделі $AP(p)$.
7. По отриманим значенням АКФ та ЧАКФ визначте q – порядок КС.

Висновки до розділу 2

Розглянуто існуючі методи прогнозу та моделювання як стаціонарних, так і нестаціонарних процесів. Виконаний аналіз їх переваг на недоліків. Метод експоненційного згладжування є самим простим, але й самим неточним, як при описі так і при прогнозуванні моделі, до того ж відсутні алгоритми для вибору оцінки або основного параметру ходу(тому він не був детально розглянутим). Чітким, теоретично обґрунтованим алгоритмом для побудови якісної моделі володіють методи АР і АРКС. Недоліками цих моделей є ускладнення моделювання їх на ЕОМ, спричинене складністю в описанні статистичних методів для вибору порядку моделі. Експертні граничні значення параметрів є гарним способом вийти з подібної ситуації. Також описано тест, що використовується для визначення наявності нестаціонарності по тренду у процесі. Розглянуто моделі для опису детермінованих трендів. Найпростішим і найпоширенішим серед яких є опис з використанням поліномів часу.

РОЗДІЛ 3. ПОБУДОВА МОДЕЛЕЙ ТА ОЦІНЮВАННЯ ПРОГНОЗІВ

3.1. Архітектура процесу побудови

Послідовність побудови моделі нестационарного процесу та оцінювання прогнозу представимо у вигляді наступної блок-схеми (рис. 3.1).

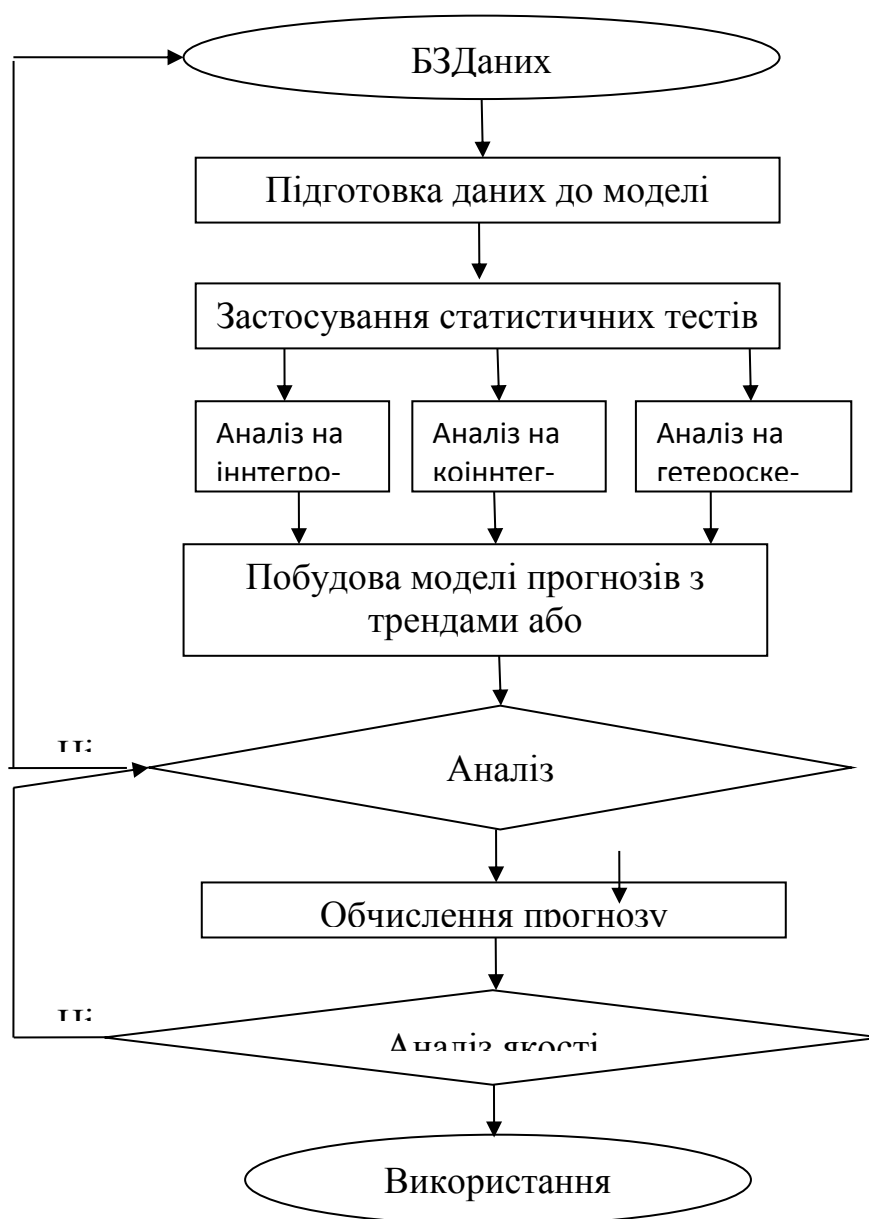


Рисунок 3.1 — Блок-схема побудови і аналізу моделі нестационарного процесу

3.2. Перевірка присутності нестационарності (тест Дікі-Фуллера)

Після виконання візуального контролю необхідно застосувати формальні тести на стаціонарність, які дають можливість переконатись в її існуванні. Розглянемо порядок застосування тесту Дікі-Фуллера.

Цей критерій допомагає визначити величину коефіцієнта a_1 в рівнянні:

$$y(k) = a_1 y(k-1) + \varepsilon(k) \quad (3.1)$$

тобто, $a=1$ чи $a<1$. Якщо $a=1$, то дані містять одиничний корінь і степінь інтегрованості дорівнює буде $I(1)$. Якщо ж $0 < a_1 < 1$, то ряд стаціонарний, тобто має степінь інтегрованості $I(0)$. Для фінансово-економічних процесів значення $a_1 > 1$ не є характерним, тому що такі значення означають присутність в процесах різко зростаючих (спадаючих) ефектів. Виникнення таких процесів є малоюмовірним, оскільки фінансово-економічне середовище є достатньо інерційним і не дозволяє змінним приймати нескінченно великі значення за короткі проміжки часу.

Нагадаємо, що застосування МНК для оцінювання коефіцієнтів моделі часового ряду передбачає, що залишки (похибки) $e(k)$ моделі мають постійну скінченну дисперсію. Присутність нестационарності приводить до порушення цього припущення. Наприклад, розглянемо рівняння

$$\begin{aligned} y(k) &= y(k-1) + e(k) = [y(k-2) - e(k-1)] + e(k) = \dots \\ &= y(0) + e(k) + e(k-1) + e(k-2) + \dots + e(1) \end{aligned} \quad (3.2)$$

Оскільки залишки $e(k)$ незалежні і мають постійну дисперсію, то дисперсія $y(k)$ зростає до нескінченності при $k \rightarrow \infty$. В такому випадку для описання динаміки ряду можна скористатись рівнянням

$$\Delta y(k) = b y(k-1) + e(k) \quad (3.3)$$

де $b = a_I - 1$.

Якщо $b = 0$, то ряд містить одиничний корінь і має степінь інтегрованості $I(1)$, а ряд $\{\Delta y(k)\}$ може бути вже стаціонарним. Якщо ж $b < 0$, то $a < 1$ і стаціонарним буде сам ряд $\{y(k)\}$.

В рівнянні $y(k) = a_I y(k-1) + \varepsilon(k)$ відсутнє середнє значення і описання тренда. Якщо включити середнє, то воно приймає вигляд:

$$y(k) = a_0 + a_I y(k-1) + \varepsilon(k) \quad (3.4)$$

або

$$\Delta y(k) = a_0 + a_I y(k-1) - y(k-1) + \varepsilon(k) = a_0 + b y(k-1) + \varepsilon(k) \quad (3.5)$$

Із врахуванням тренда останнє рівняння приймає вигляд:

$$y(k) = a_0 + a_I k + a_2 y(k-1) + \varepsilon(k) \quad (3.6)$$

де k – дискретний час. Це рівняння можна записати для першої різниці

$$y(k) - y(k-1) = a_0 + a_I k + b y(k-1) - y(k-1) + \varepsilon(k) \quad (3.7)$$

або

$$\Delta y(k-1) = a_0 + a_1 k + b y(k-1) + \varepsilon(k) \quad (3.8)$$

Для такої моделі було б некоректно використовувати t -статистику з метою визначення значимості коефіцієнта b , оскільки застосування регресії для оцінювання цього коефіцієнта передбачає, що $b < 0$ ($a_1 < 1$). Тобто при $b \approx 0$ великий процент оцінок за t -статистикою не буде прийматися як значимий, тобто нульова гіпотеза щодо існування одиничного кореня буде часто відкидатись.

Крім того, одиничні корені робастні (зберігаються і можуть бути виявлені) при різних степенях гетероскедастичності, але можуть виникати проблеми з автокореляцією залишків моделі. В умовах присутності автокореляції залишків задача тестування на стаціонарність розв'язується за допомогою розширеного тесту Дікі-Фуллера. При використанні цього методу значення залежної змінної вводяться в рівняння регресії з великими значеннями лагу, достатніми для того щоб уникнути автокореляції залишків. Це рівняння може мати наступний вигляд:

$$\Delta y(k) = a_0 + b y(k-1) + c_1 \Delta y(k-1) + c_2 \Delta y(k-2) + \dots + c_n y(k-n) + \varepsilon(k) \quad (3.9)$$

Форма критерію значимості залежить від виду моделі, що тестується, тобто, чи включено в модель середнє значення і член, який описує тренд.

Нульова гіпотеза без середнього.

При тестуванні рівняння

$$\Delta y(k) = b y(k-1) + c_1 \Delta y(k-1) + c_2 \Delta y(k-2) + \dots + c_n y(k-n) + \varepsilon(k) \quad (3.10)$$

тобто, середнє відсутнє, гіпотеза записується так:

$H_0 : b = 0$ — ряд нестационарний;

$H_1 : b < 0$ — ряд стаціонарний.

Нульова гіпотеза відкидається, якщо статистика b/SE_b має від'ємне значення, яке менше за критичне значення, взяте з таблиці Дікі-Фуллера. Критичні значення для рівнів значимості $\alpha=1$ і $\alpha=5$ дорівнюють $-2,58$ і $-1,95$, відповідно.

Якщо нульова гіпотеза приймається, то ряд $\{y(k)\}$ - це випадкове блукання без зсуву (константи в рівнянні).

В більш загальному вигляді цього критерію враховується розмір вибірки N , що досягається шляхом обчислення модифікованого критичного значення по формулі

$$\tau_{\infty} + \frac{\tau_1}{N} + \frac{\tau_2}{N^2} \quad (3.11)$$

де

$$\tau_{\infty} = -2,57 \ (\alpha=1) \text{ або } \tau_{\infty} = -1,94 \ (\alpha=5);$$

$$\tau_1 = -1,96 \ (\alpha=1) \text{ або } \tau_1 = -0,398 \ (\alpha=5);$$

$$\tau_2 = -10,04 \ (\alpha=1) \text{ або } \tau_2 = 0 \ (\alpha=5);$$

(значення τ табульовані Маккінномом, 1991).

Нульова гіпотеза з середнім значенням

Перевірка рівняння $\Delta y(k) = a_0 + b y(k-1) + e(k)$ із врахуванням можливої автокореляції залишків (як це було показано вище) базується на використанні того ж статистичного критерію, що і для рівняння без середнього, і тієї ж формули критичних значень, але при наступних значеннях τ :

$$\tau_{\infty} = -3,43 \ (\alpha=1) \text{ або } \tau_{\infty} = -2,86 \ (\alpha=5);$$

$$\tau_1 = -6,00 \ (\alpha=1) \text{ або } \tau_1 = -2,74 \ (\alpha=5);$$

$$\tau_2 = -29,25 \ (\alpha=1) \text{ або } \tau_2 = -8,36 \ (\alpha=5).$$

Нульова гіпотеза при наявності середнього та тренду.

В даному випадку застосовується така ж процедура, що і вище, але при наступних значеннях τ :

$$\tau_{\infty} = -3,96 \ (\alpha=1) \text{ або } \tau_{\infty} = -3,41 \ (\alpha=5);$$

$$\tau_1 = -8,35 \ (\alpha=1) \text{ або } \tau_1 = -4,04 \ (\alpha=5);$$

$$\tau_2 = -47,44 \ (\alpha=1) \text{ або } \tau_2 = -17,83 \ (\alpha=5).$$

3.3. Розширений тест Дікі-Фуллера

Для того щоб скористатись тестом ДФ, необхідно побудувати наступне рівняння регресії:

$$\Delta y(k) = a_0 + a_1 k + b y(k-1) + \sum_{i=1}^p c_i \Delta y(k-i) + \varepsilon(k) \quad (3.12)$$

де a_0, a_1, b, c_i – невідомі коефіцієнти регресії. Якщо всі коефіцієнти $c_i = 0, i = 1, 2, \dots, p$, то рівнянням (3.12) можна скористатись для застосування тесту ДФ, інакше необхідно використати розширений тест ДФ. На практиці рекомендують застосовувати тест РДФ з кількістю затриманих у часі значень p меншою 10% числа спостережень, тобто $p < 0,1 N$, де N – довжина (потужність) часового ряду. При використанні тесту ДФ і РДФ важливо правильно задати структуру моделі, зокрема, необхідно визначити чи потрібно включати члени a_0 і $a_1 k$.

Пропонується наступне евристичне правило: якщо візуально з графіка не можна зробити висновок про наявність тренду, то в модель (3.9) необхідно

включати тільки константу (перетин) a_0 , навіть якщо значення коливаються навколо нуля. Якщо візуальний аналіз ряду свідчить про присутність тренду, то в модель (3.12) необхідно ввести a_0 і $a_1 k$.

За допомогою базової моделі (3.12) тестуються такі гіпотези:

$H_0 : b = 0$, або часовий ряд нестационарний: $\{y(k)\} \square I(\text{int}), \text{int} > 0$;

$H_1 : b < 0$, або часовий ряд стаціонарний: $\{y(k)\} \square I(0), \text{int} = 0$.

Нульова гіпотеза відкидається, якщо отримана оцінка коефіцієнта $\hat{b} < 0$ та обчислена τ – статистика Маккіннона (для тестування на наявність одиничного кореня) за абсолютною величиною більша за величину критичного значення цієї статистики при вибраному рівні значимості α . Формально це можна записати так:

$$|\tau| = \left| \frac{\hat{b}}{SE_{\hat{b}}} \right| \geq |\tau_{\text{крит}}| \quad (3.13)$$

на рівні значимості α ; $SE_{\hat{b}}$ – стандартна похибка оцінки \hat{b} .

3.4. Перевірка процесу на стаціонарність (наявність тренду) за допомогою тесту Дікі–Фуллера

При визначенні наявності нестационарності (тобто присутності одиничного кореня), пропонується скористатися тестом Дікі–Фуллера [9], суть якого полягає в наступному: для визначення присутності одиничного кореня запропоновано скористатись трьома наступними рівняннями:

$$\Delta y(k) = \gamma y(k-1) + \varepsilon(k) \quad (3.14)$$

$$\Delta y(k) = a_0 + \gamma y(k-1) + \varepsilon(k) \quad (3.15)$$

$$\Delta y(k) = a_0 + \gamma y(k-1) + a_2 k + \varepsilon(k) \quad (3.16)$$

де k – представляє собою дискретний час; $\gamma = a_1 - 1$ – є коефіцієнтом рівняння: $y(k) = a_0 + a_1 y(k-1) + \varepsilon(k)$. Присутність детермінованих членів a_0 і $a_2 k$ у рівняннях (3.15) і (3.16) являється різницею між рівняннями (3.14) та (3.15), (3.16). Блуканням називають модель випадкового кроку. Рівняння (3.14) представляє собою блукання, друге рівняння у вигляді константи a_0 включає в себе зсув, третє ж рівняння включає в собі детермінований лінійний часовий тренд та зсув.

У розглянутих рівняннях інтерес для нас представляє параметр γ . Послідовність $\{y(k)\}$ має одиничний корінь, якщо $\gamma = 0$. Тест Дікі-Фуллера надалі будемо обзначати, як ДФТ. Застосовуючи ДФТ ми, за допомогою МНГ або ММП, маємо змогу провести оцінювання наведених вище рівнянь з метою отримання стандартної похибки отриманої оцінки та оцінки параметра γ . t – статистика обчислюється на основі оцінки та її стандартної похибки, яку надалі потрібно порівняти із наведеним в таблицях значеннями. Рішення про справедливість або відхилення нуль-гіпотези ($\gamma = 0$), приймається на основі цього порівняння.

Нехай проведено 100 спостережень, тоді на основі цих спостережень отримаємо оцінку параметра $\hat{a}_1 = 0,9459$ для рівняння $y(k) = a_1 y(k-1) + \varepsilon(k)$ із стандартною похибкою $SE_{a_1} = 0,031$. Зрозуміло, що виконавши оцінювання рівняння $\Delta y(k) = \gamma y(k-1) + \varepsilon(k)$ ми отримаємо оцінку $\hat{\gamma} = -0,0541$ з тією ж стандартною похибкою 0,031. Таким чином підрахувавши t – статистику для нуль-гіпотези $\gamma_0 = 0$, отримаємо $t = -0,0541/0,031 = -1,74516$.

Для випадку $a_0 = a_2 = 0$ при $N = 100$, застосувавши таблиці Дікі-Фуллера знаходимо критичні значення t – статистики, що при рівнях значимості 10%, 5% та 1% мають наступні значення: -1,61; -1,95 та -2,60, відповідно. Таким

чином, в розглянутому гіпотетичному випадку при $\hat{\gamma} = -0,0541$ нуль-гіпотеза відхиляється при рівні значимості 10%, але не може бути відхилена при рівнях значимості 5% та 1%. Спостерігаємо зменшення критичних значень t -статистики при збільшенні розміру вибірки для будь-якого рівня значимості.

Тестування на присутність одиничного кореня рівнянь (3.14) – (3.16) проводимо за тією ж методикою. Однак, критичні значення t -статистики залежать від того чи присутності зсуву α_0 та детермінованого тренду $\alpha_2 k$ в структурі моделі. Від структури регресійного рівняння та від довжини вибірки залежать критичні значення для $\gamma = 0$. Так, для рівняння (3.14) застосовуємо статистику позначену через τ , для рівняння (3.15) – статистика позначається τ_μ , а для рівняння (3.16) – через τ_τ .

При включенні в модель константи(зсуву), при $a_2 = 0$, для отримання критичних значень t -статистики необхідно користуватись іншою частиною таблиці. Для гіпотетичного прикладу зробимо оцінку рівняння у формі $\Delta y(k) = a_0 + \gamma y(k-1) + \varepsilon(k)$. Отримаємо наступні результати : $\gamma = 0,9135 - 1 = -0,0865$ із стандартною похибкою 0,041. Тепер для t -статистики отримаємо наступне значення : $\tau_\mu = -0,0865/0,041 = -2,1098$.

Використавши таблиці, знаходимо величину критичних значень -2,58; -2,89 та -3,51 на рівнях значимості 10%, 5% та 1%, відповідно для 100 спостережень. Робимо висновок про присутність одиничного кореня і про неможливість відхилення нуль-гіпотези при всіх рівнях значимості, закладених в таблицю Дікі-Фуллера.

При використанні структури моделі іншого вигляду : $\Delta y(k) = a_0 + a_1 y(k-1) + a_2 k + \varepsilon(k)$, отримаємо наступні значення статистики τ_τ - 3,45 та - 4,04 при рівнях значимості 5% і 1%, відповідно.

Критичні значення не зміняться, якщо рівняння (3.14), (3.15) і (3.16) замінити наступними рівняннями авторегресії:

$$\Delta y(k) = \gamma y(k-1) + \sum_{i=2}^p \beta_i \Delta y(k-i+1) + \varepsilon(k) \quad (3.17)$$

$$\Delta y(k) = a_0 + \gamma y(k-1) + \sum_{i=2}^p \beta_i \Delta y(k-i+1) + \varepsilon(k)$$

$$\Delta y(k) = a_0 + \gamma y(k-1) + \sum_{i=2}^p \beta_i \Delta y(k-i+1) + \varepsilon(k)$$

$$\Delta y(k) = a_0 + \gamma y(k-1) + a_2 k + \sum_{i=2}^p \beta_i \Delta y(k-i+1) + \varepsilon(k)$$

Тобто ті ж статистики: τ , τ_μ і τ_τ використовуються при проведенні тестів на нуль-гіпотезу. F – статистики: ϕ_1 , ϕ_2 і ϕ_3 запропоновані для перевірки об'єднаних гіпотез щодо коефіцієнтів Дікі і Фулер.

Так, статистика ϕ_1 застосовується при перевірці нуль-гіпотези щодо $\gamma = a_0 = 0$ в рівняннях (3.4.2) і (3.4.5). Статистику ϕ_2 застосовуємо при перевірці об'єднаної гіпотези $a_0 = \gamma = a_2 = 0$ в рівняннях (3.4.3) і (3.4.6), а статистика ϕ_3 – для перевірки об'єднаної гіпотези $\gamma = a_2 = 0$.

Статистики ϕ_1 , ϕ_2 і ϕ_3 обчислюються по аналогії із звичайними F – статистиками: $\phi_i = \frac{[RSS_1 - RSS_2]r}{RSS_2(N-n)}$, де RSS_1 і RSS_2 – суми квадратів похибок (СКП), обчислених для моделей з обмеженнями та моделей без обмежень; r – число обмежень; N – число використаних спостережень; n – число параметрів, оцінених для необмеженої моделі. Можливими обмеженнями можуть бути обмеження на порядок моделі та їх структуру.

Для визначення рівня значимості, при якому обмеження моделі будуть відігравати суттєву роль потрібно обчислити значення ϕ_i та отримати його

порівняння із відповідним значенням Дікі і Фулера. За нуль-гіпотезу приймають генерування даних моделі з обмеженням. Генерацію даних моделі без обмежень приймають за альтернативу. Сума квадратів похибок для моделі з обмеженням буде приближеним до СКП для моделі без обмежень у випадку несуттєвої ролі обмежень. Очевидно, в такому випадку, ϕ_i приймає невелике значення, а при відігриванні обмеженням суттєвої ролі і відхиленні нуль-гіпотези, ϕ_i матиме великі значення.

Для визначення рівня значимості, при якому обмеження моделі будуть відігравати суттєву роль потрібно обчислити значення ϕ_i та отримати його порівняння із відповідним значенням Дікі і Фулера. За нуль-гіпотезу приймають генерування даних моделі з обмеженням. Генерацію даних моделі без обмежень приймають за альтернативу. Сума квадратів похибок для моделі з обмеженням буде приближеним до СКП для моделі без обмежень у випадку несуттєвої ролі обмежень. Очевидно, в такому випадку, ϕ_i приймає невелике значення, а при відігриванні обмеженням суттєвої ролі і відхиленні нуль-гіпотези, ϕ_i матиме великі значення.

Необхідно також виконати перевірку гіпотез щодо значимості константи a_0 та коефіцієнта a_2 . Значимість коефіцієнта a_2 . Свідчить про присутність часового тренду. У випадку визначення нуль-гіпотези як $\gamma = 0$, для тестування на присутність часового тренду використаємо статистики $\tau_{\beta\tau}$. Отже, данною статистикою виконується перевірка на значення $a_2 = 0$ при $\gamma = 0$. Для перевірки гіпотези щодо $a_0 = 0$, використовується статистика $\tau_{\alpha\tau}$ якщо оцінюється модель (3.19), або статистикою $\tau_{\alpha\mu}$, якщо оцінюється модель (3.18).

3.5 Критерії адекватності моделі

Критерії адекватності моделі дозволяють оцінити окремо значимість коефіцієнтів математичної моделі в статистичному сенсі, визначити інтегральну похибку моделі стосовно вихідного часового ряду, встановити наявність кореляції між значеннями помилки моделі (нагадуємо, що вони повинні бути не корельованими), а також визначити ступінь адекватності моделі фізичному процесу в цілому. У цю множину входять такі статистичні параметри:

1. t – статистика Стюдента. Для визначення кожного коефіцієнта регресії в статистичному сенсі використовують t – статистику, що в більшості статистичних пакетів оцінюється за наступною формулою:

$$t_a = \frac{\hat{a} - a_0}{SE_a} \quad (3.21)$$

де \hat{a} – оцінка коефіцієнту; a_0 – нуль гіпотеза у відношенні значення цього коефіцієнту (звичайно $a_0 = 0$); SE_a – стандартна похибка оцінки коефіцієнта, що обчислюється пакетом. Очевидно, що чим менше значення стандартної похибки, тим кращою є оцінка коефіцієнта для моделі.

Щоб визначити значимість коефіцієнта знати наступні значення: довжину вибірки N , число оцінюваних параметрів p і задати рівень значимості α (у звичайних випадках задаються $\alpha = 1\%$, $\alpha = 5\%$ або $\alpha = 10\%$) Рівень значимості означає, процент випадків у якому ми допускаємо помилкове прийняття рішень про значимість оцінок. Для вибору цих параметрів використовують таблиці значення $t_{крит}$. При $-t_{крит} < t_a < t_{крит}$, приймається нуль-гіпотеза не значимості коефіцієнта. У протилежному випадку коефіцієнт вважається значимим, так як відбудеться відхилення цієї нуль-гіпотези. Значення статистики t_a обернено пропорційна стандартній похибці SE_a . Таким чином

збільшення значення t_a , буде означати збільшення значимості конкретного коефіцієнта

2. Часто, за міру інформативності часового ряду приймають дисперсію коефіцієнта детермінації R^2 . Коефіцієнт R^2 - це відношення дисперсії тієї частини часового ряду основної змінної, що описується отриманим рівнянням, до вибіркової дисперсії цієї змінної[3]. Обчислюється наступною формулою :

$$R^2 = \frac{\text{var}(\hat{y})}{\text{var}(y)} \quad (3.22)$$

Очевидно, що для адекватної моделі коефіцієнт детермінації повинен прямувати до одиниці, тобто: $R^2 \rightarrow 1$.

3. Сума квадратів помилок моделі $\sum e^2(k)$, тобто

$$SSE = \sum_{k=1}^N [\hat{y}(k) - y(k)]^2 \quad (3.23)$$

де

$$\hat{y}(k) = \hat{a}_0 + \hat{a}_1 \hat{y}(k-1) + \hat{a}_2 \hat{y}(k-2) + \hat{b}_1 x(k) + \hat{b}_2 z(k) \quad (3.24)$$

де $y(k)$ – вимірювання; N – довжина вибірки. З усіх моделей-кандидатів потрібно обрати ту модель, для якої $\sum e^2(k)$ приймає мінімальне значення.

4. Інформаційний критерій Акайке (AIC). Цей критерій враховує суму квадратів похибок, число вимірювань N і число оцінюваних параметрів p :

$$AIC = N \ln \left[\sum_{k=1}^N e^2(k) \right] + 2p. \text{ Очевидно, що для кращої моделі критерій має}$$

менше значення, оскільки він залежить від суми квадратів похибок (СКП).

Проте, крім СКП, даний критерій враховує довжину вибірки і число оцінюваних параметрів, що робить його більш інформативним[4].

5. Статистика Дарбіна-Уотсона(Durbin-Watson). Статистика Дарбіна-Уотсона обчислюється за формулою: $DW = 2 - 2\rho$, де ρ – коефіцієнт кореляції між значеннями випадкової змінної $\varepsilon(k) \approx e(k)$, тобто $\rho = \text{cov}[e(k)] = E[e(k)e(k-1)]$. Цей параметр дозволяє визначити ступінь корельованості похибок моделі. При повній відсутності кореляції між похибками $DW = 2$, це найбільше прийнятне значення даного параметра[4].

6. Статистика Фішера F , яка визначає ступінь адекватності моделі в цілому. Для адекватної моделі виконується умова: $F > F_{\text{крит}}$, де $F_{\text{крит}}$ визначається по таблиці аналогічно t – статистиці. Значення F пропорційно $R^2 / (1 - R^2)$, де R^2 – коефіцієнт детермінації. Таким чином, більшому значенню F відповідає більш адекватна модель[4].

7. Коефіцієнт Тейлає дуже важливим індикатором точності моделі і її сумісності [4]:

$$U = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i)^2 + \frac{1}{N} \sum_{i=1}^N (\hat{y}_i)^2}} \quad (3.25)$$

За побудовою, його величина знаходиться між 0 і 1. Якщо $U=1$, модель не може бути використана для прогнозу. Прогнозовані, на основі отриманої моделі, і реальні ряди є не корельованими. У протилежному випадку, якщо $U=0$, прогнозовані ряди співпадають з реальними рядами і модель є ідеальною.

Цей коефіцієнт може бути розкладений на суму відношення U^M , відношення варіацій U^S і відношення коваріацій U^C .

$$U^M = \frac{(y_i - \hat{y}_i)^2}{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (3.26)$$

і використовується для перевірки наявності систематичних відхилень для середніх реальних і прогнозованих рядів. Або, інакше кажучи, чи модель ввесь час завищує прогноз. Чим менше величина U^M тим краще. Якщо $U^M = 0$, то в прогнозованих значеннях відсутня упередженість і модель є якісною.

U^S – відношення варіацій і визначається як:

$$U^M = \frac{(\sigma_{actual} - \sigma_{fitted})^2}{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (3.27)$$

Відношення варіацій використовується для перевірки того, що модель має достатньо динамічних властивостей для поглинання варіацій реальних рядів. Наприклад, модель може забезпечити систематично менші коливання, ніж коливання реальних рядів. Аналогічно U^M , менші значення U^S є індикатором меншого зміщення.

U^C – відношення коваріацій і визначається як:

$$U^C = \frac{2(1 - \rho)(\sigma_{actual} - \sigma_{fitted})^2}{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (3.28)$$

Відношення коваріацій U^C вимірює, степінь корельованості прогнозованих і реальних рядів. Рівність U^C нулює свідченням того, що прогнозовані і реальні ряди ідеально корельовано. Необхідно зазначити, що $U^C + U^S + U^M = 1$

MAPE - середня абсолютна помилка у відсотках використовується для оцінки точності прогнозу; показує на скільки великі помилки в порівнянні зі

значеннями ряду. Розраховується як [6]:

$$MAPE = \frac{1}{n} \sum_{i=1}^N \frac{y_i - \hat{y}_i}{y_i} \quad (3.29)$$

Коли нам потрібно підвгнати криву під наші данні, то точність цієї підгонки буде оцінена програмою за середньоквадратичною похибкою(середня квадратична похибка, MSE).

Насправді, програма, розповсюджуючи криву підгонки, прагне мінімізувати цей коефіцієнт. Фізичного смисла MSE не має, але чем ближче до нуля, тим модель краща.

Середньоквадратична похибка (СКП):

$$СКП = \sqrt{\frac{1}{S} \sum_{i=1}^S (y(k+s) - \hat{y}(k+s, k))^2} \quad (3.30)$$

3.6. Побудова моделей вибраних процесів за статистичними даними

NullHypothesis: CLOSE has a unitroot

Exogenous: Constant

LagLength: 0 (Automatic - basedon SIC, maxlag=12)

		t-Statistic	Prob.*
AugmentedDickey-Fullerteststatistic		-1.135256	0.7004
Testcriticalvalues:	1% level	-3.482879	
	5% level	-2.884477	
	10% level	-2.579080	

*MacKinnon (1996) one-sided p-values.

AugmentedDickey-FullerTestEquation

DependentVariable: D(CLOSE)

Method: LeastSquares

Date: 12/19/19 Time: 16:23

Sample (adjusted): 6/04/2019 11/26/2019

Includedobservations: 126 afteradjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CLOSE(-1)	-0.018196	0.016028	-1.135256	0.2585
C	0.798749	0.808152	0.988364	0.3249
R-squared	0.010287	Meandependentvar		-0.115794
Adjusted R-squared	0.002305	S.D. dependentvar		0.723632
S.E. of regression	0.722798	Akaikeinfocriterion		2.204372
Sumsquaredresid	64.78217	Schwarzcriterion		2.249392
Loglikelihood	-136.8754	Hannan-Quinn criter.		2.222662
F-statistic	1.288806	Durbin-Watson stat		1.753770
Prob(F-statistic)	0.258458			

Кореляція в в заданій моделі представлена на рис. 3.2.

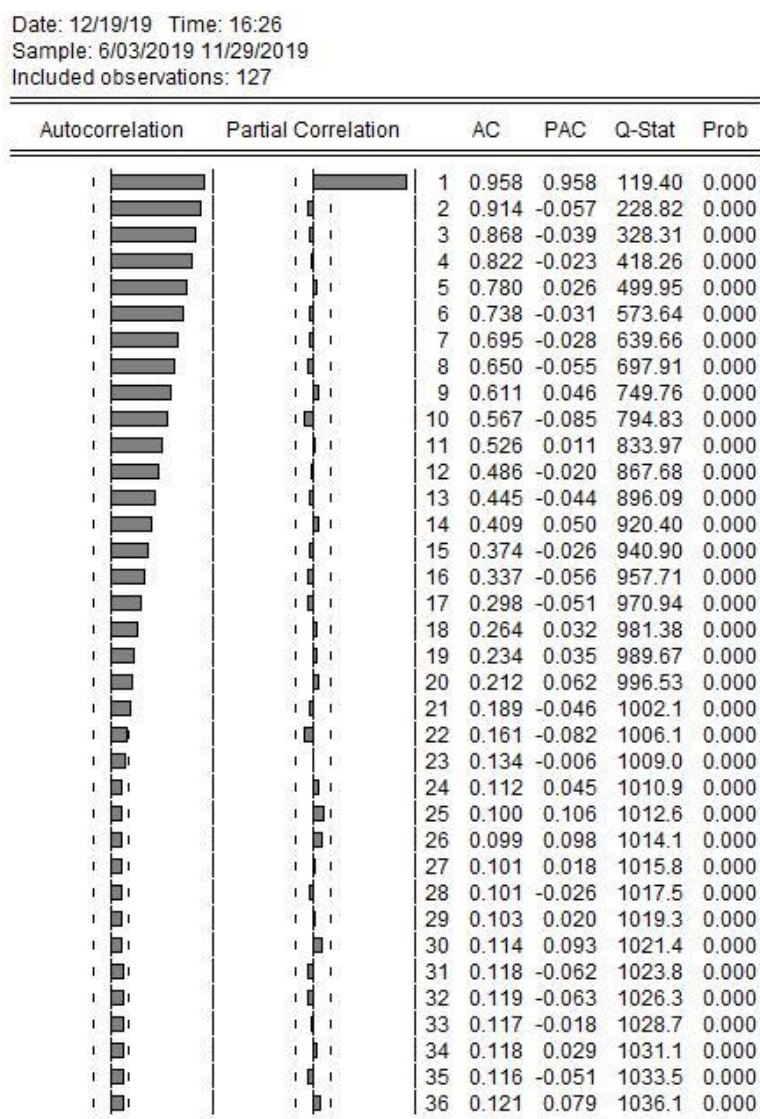


Рисунок 3.2 — Кореляція в заданій моделі

Оцінка рівняння представлена наступними даними:

Dependent Variable: CLOSE

Method: LeastSquares (Gauss-Newton / Marquardtsteps)

Date: 12/19/19 Time: 16:34

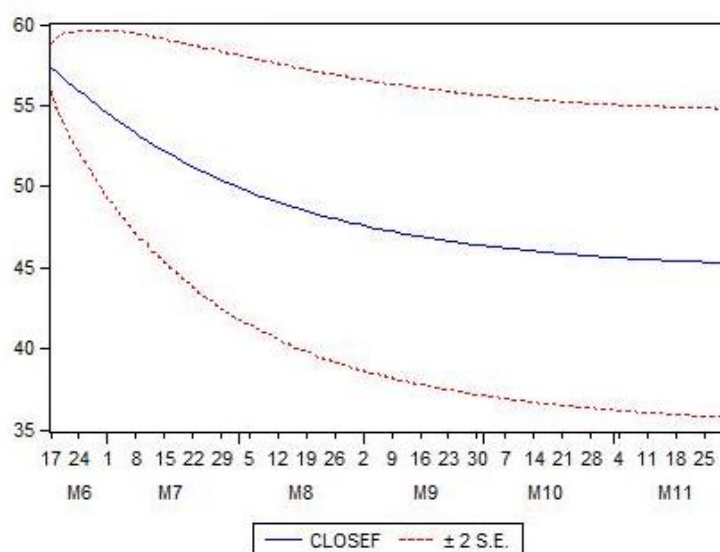
Sample (adjusted): 6/17/2019 11/26/2019

Included observations: 117 after adjustments

$CLOSE = C(1) + C(1) * CLOSE(-1) + C(2) * CLOSE(-10)$

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.996338	0.026979	36.93023	0.0000
C(2)	-0.018567	0.026992	-0.687862	0.4929
R-squared	0.957076	Meandependentvar		49.56061
Adjusted R-squared	0.956702	S.D. dependentvar		3.555514
S.E. of regression	0.739834	Akaikeinfocriterion		2.252164
Sumsquaredresid	62.94570	Schwarzcriterion		2.299380
Loglikelihood	-129.7516	Hannan-Quinn criter.		2.271333
Durbin-Watson stat	1.797443			

Графік обчислення прогнозу представлений на рис.3.3.



Forecast: CLOSEF
 Actual: CLOSE
 Forecast sample: 6/03/2019 11/29/2019
 Adjusted sample: 6/17/2019 11/29/2019
 Included observations: 117
 Root Mean Squared Error 2.085566
 Mean Absolute Error 1.671238
 Mean Abs. Percent Error 3.360997
 Theil Inequality Coefficient 0.021184
 Bias Proportion 0.190417
 Variance Proportion 0.010921
 Covariance Proportion 0.798662

Рисунок 3.3 — Обчислення прогнозу

3.7. Дослідження та моделювання ряду Лоренца

Розглянемо роботу нашої СППР при аналізі та моделюванні тестового ряду Лоренца (рис. 3.4, 3.6).

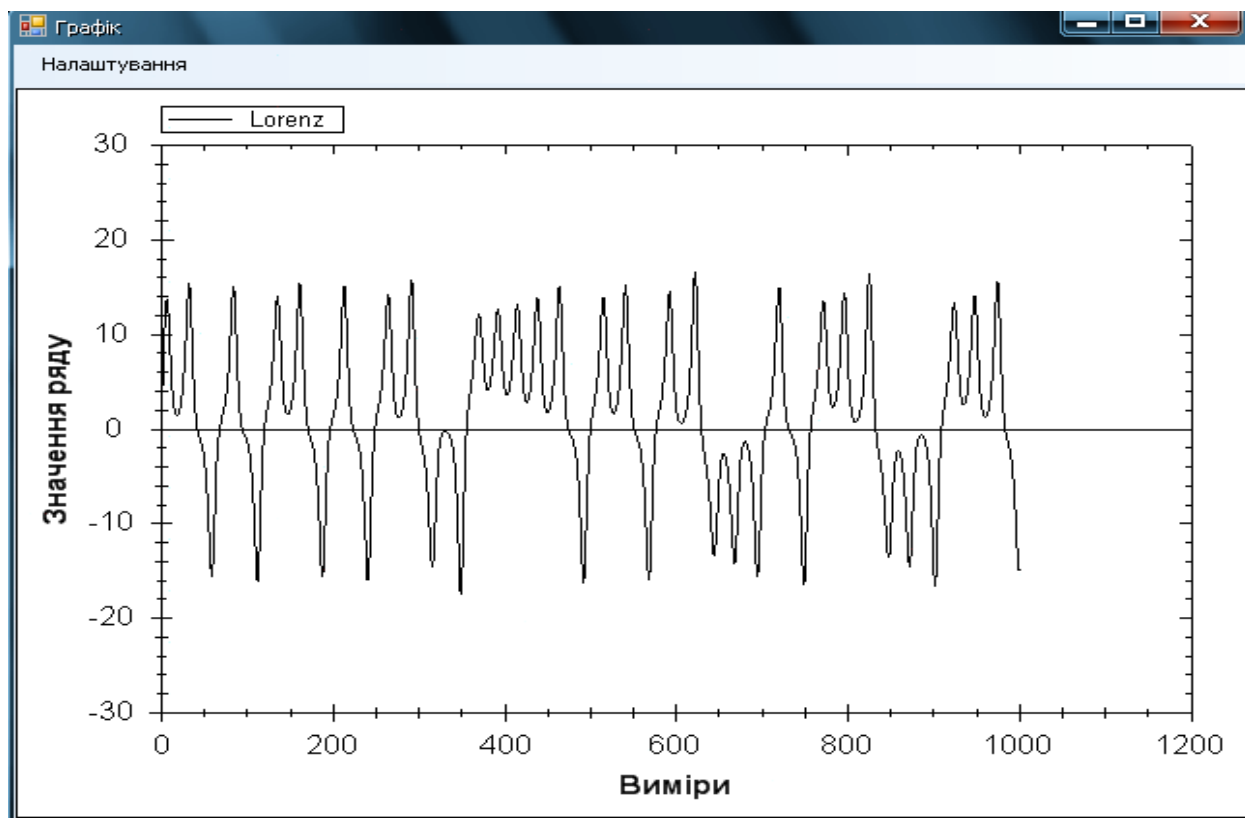


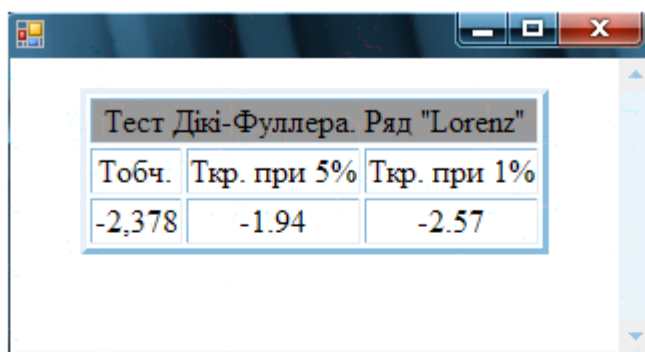
Рисунок 3.4 — Графічне представлення ряду Лоренца

Результати тестування на нестационарність підтверджують нашу здогадку про те, що ряд стаціонарний.

Судячи із АКФ та ЧАКФ (рис.3.3) буде доцільно побудувати наступні моделі АР п'ятого порядку без врахування нульового коефіцієнту:

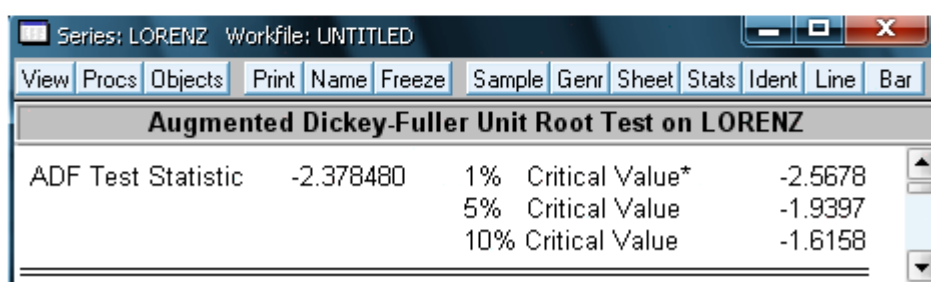
$$1) y(k) = a_1 * y(k-1) + a_2 * y(k-2) + a_4 * y(k-4) + a_5 * y(k-5) + e(k)$$

$$2) y(k) = a_1 * y(k-1) + a_2 * y(k-2) + a_3 * y(k-3) + a_4 * y(k-4) + a_5 * y(k-5) + e(k)$$



Тест Дікі-Фуллера. Ряд "Lorenz"		
Тобч.	Ткр. при 5%	Ткр. при 1%
-2,378	-1.94	-2.57

а) СППР



Augmented Dickey-Fuller Unit Root Test on LORENZ			
ADF Test Statistic	-2.378480	1% Critical Value*	-2.5678
		5% Critical Value	-1.9397
		10% Critical Value	-1.6158

б) Eviews

Рисунок 3.5 — Тест на нестационарність для ряду Лоренца

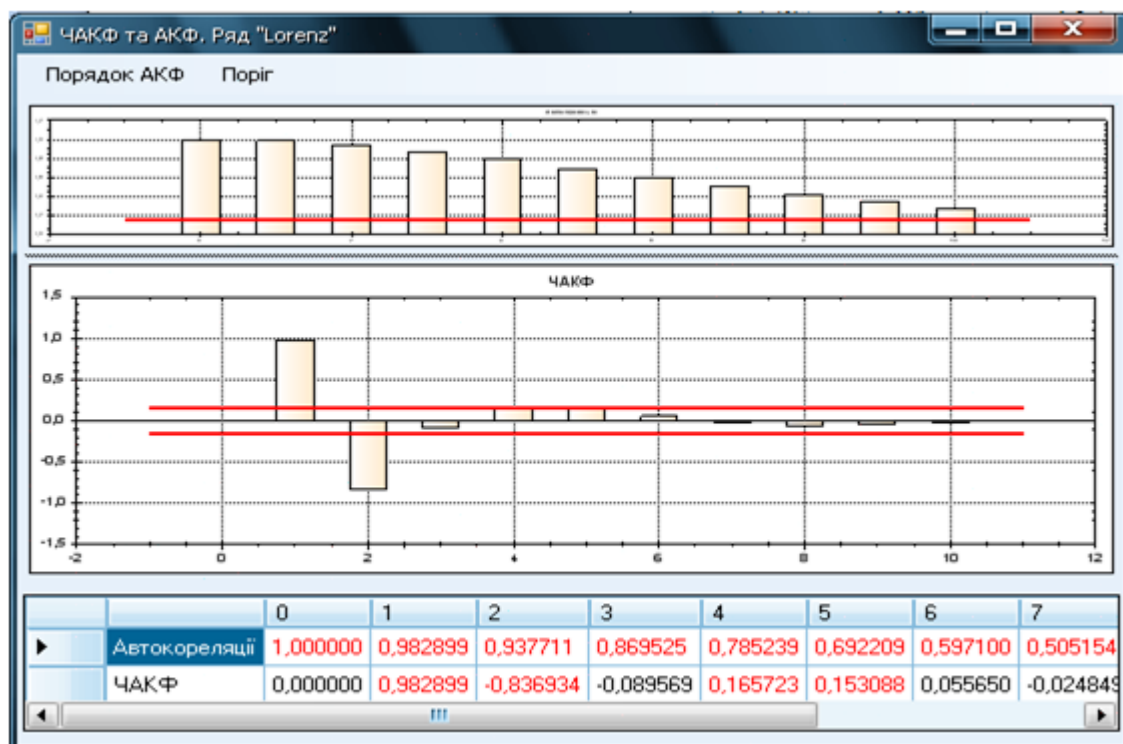
Розрахуємо коефіцієнти обох рівнянь і оцінимо АКФ та ЧАКФ залишків моделей, щоб визначитись чи варто включати до них КС. В обох випадках після включення коефіцієнтів КС отримали їх оцінки близькими до нуля, тому можемо вважати їх не значимими. Для визначення кращої моделі скористаємося розрахованими параметрами якості моделі та прогнозу. Опираючись на отримані результати вибираємо другу модель, так як у неї кращі характеристики. Результати оцінки параметрів вибраної моделі подані в таблиці 3.1. У таблиці 3.2 представлені результати прогнозування.

Таблиця 3.1. Оцінки параметрів

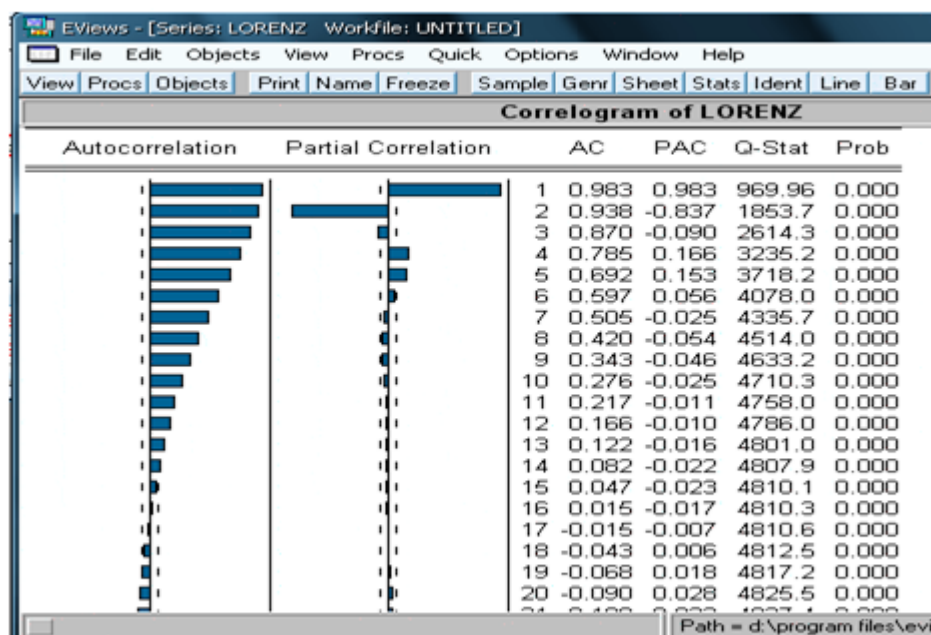
Коефіцієнт	Оцінка в СППР, метод МНК	Оцінка в СППР, метод РМНК	Оцінка в Eviews
a_1	4,2602	3,5853	4,2602
a_2	-7,6513	-5,1601	-7,6513
a_3	7,2667	3,6556	7,2667
a_4	-3,6543	-1,2157	-3,6542
a_5	0,7776	0,1309	0,7775

Таблиця 3.2. Результати прогнозування

Вимір	Реальне значення	СППР				Eviews	
		МНК		РМНК		Прогноз	Похибка
		Прогноз	Похибка	Прогноз	Похибка		
997	-12,413	-12,435	0,022	- 12,395	-0,018	- 12,437	0,024
998	-14,058	-14,129	0,071	- 14,135	0,077	- 14,131	0,073
999	-14,974	-15,059	0,085	- 15,119	0,145	- 15,061	0,087
1000	-14,784	-14,807	0,023	- 14,894	0,110	- 14,808	0,024
1001		-13,361		- 13,403		- 13,362	
Сер. Похиб- ка			0,040		0,063		0,042



а) СППР



б) Eviews

Рисунок 3.6 — АКФ та ЧАКФ для ряду Лоренца

Характеристики якості моделей для ряду Лоренца представлені на рис. 3.7.

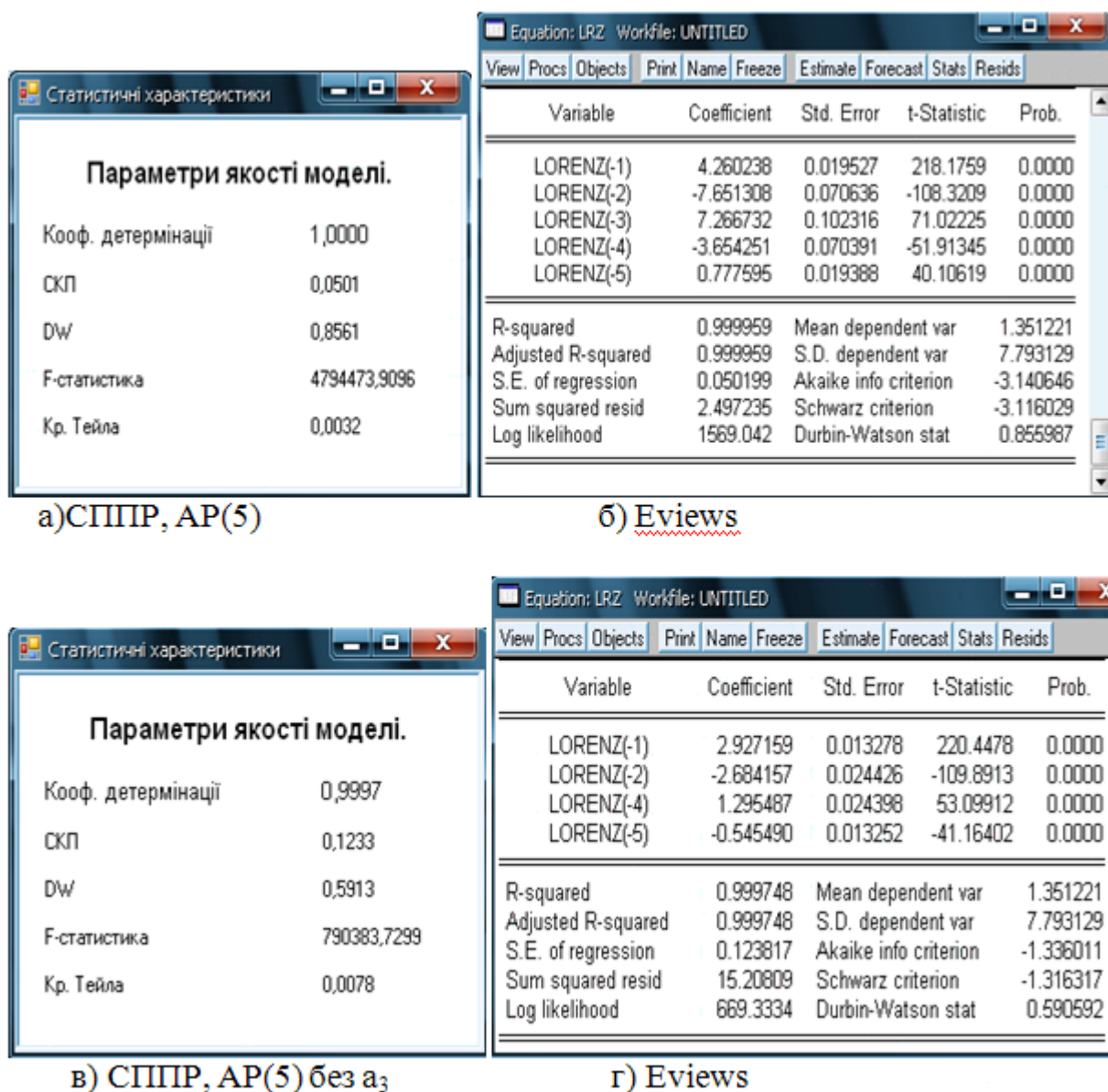


Рисунок 3.7 — Характеристики моделей для ряда Лоренца

3.8. Дослідження та моделювання ВВП України

Дослідимо поквартальні показники ВВП України, на основі значень яких побудуємо авторегресійну модель і зробимо однокроковий статичний прогноз на майбутнє.

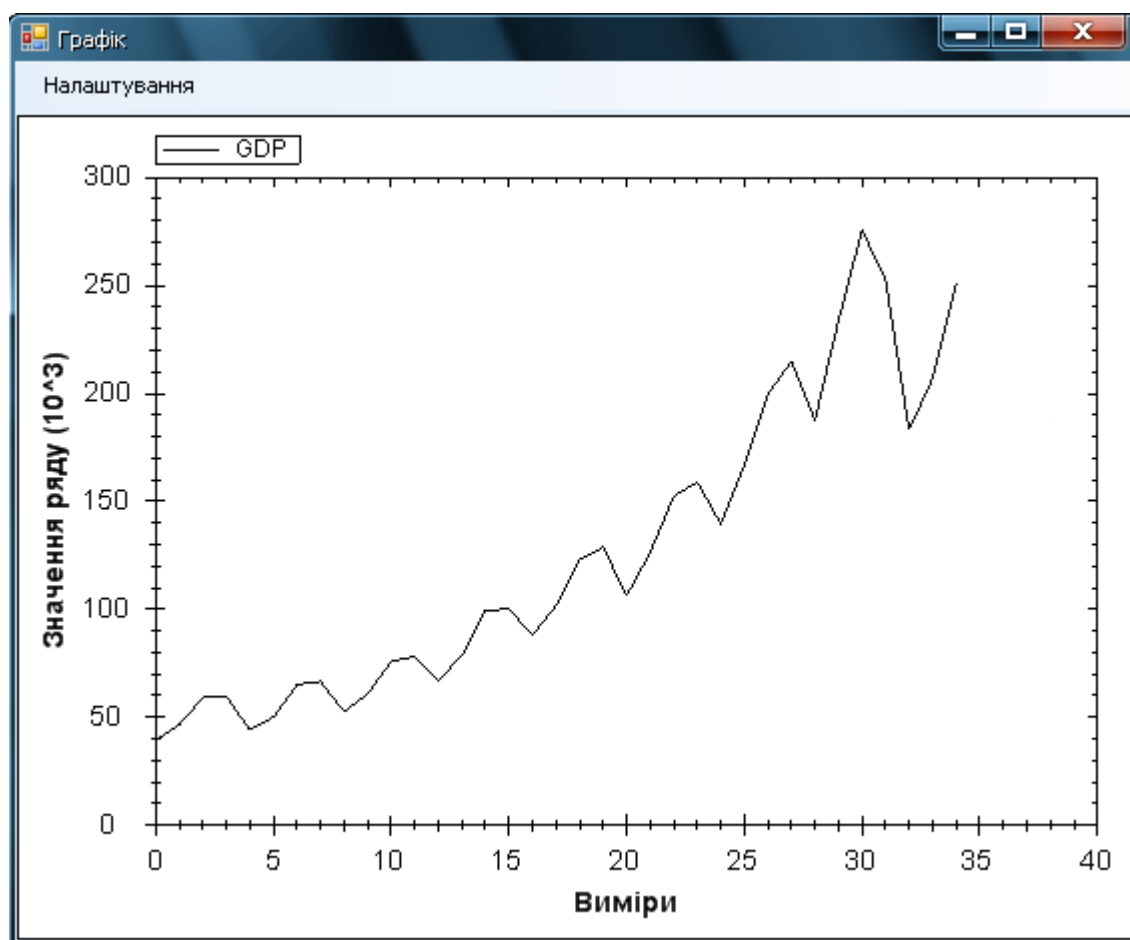


Рисунок 3.8 — Поквартальні показники ВВП України, починаючи із 2001 року, млн грн

Із графічного представлення даних можемо зробити припущення про наявність тренду (рис. 3.8). Провівши тест на нестационарність, переконуємось, що наша гіпотеза є вірною і перед подальшими дослідженнями та моделюванням варто виконати попередню обробку даних, щоб привести ряд до стаціонарного виду (рис. 3.9).

Тест Дікі-Фуллера. Ряд "GDP"		
Тобч.	Ткр. при 5%	Ткр. при 1%
1,111	-1.94	-2.57

Рисунок 3.9 — Результати тесту на нестационарність

Візьмемо перші різниці і отримаємо стаціонарний ряд, з яким будемо працювати в подальшому (рис. 3.10)

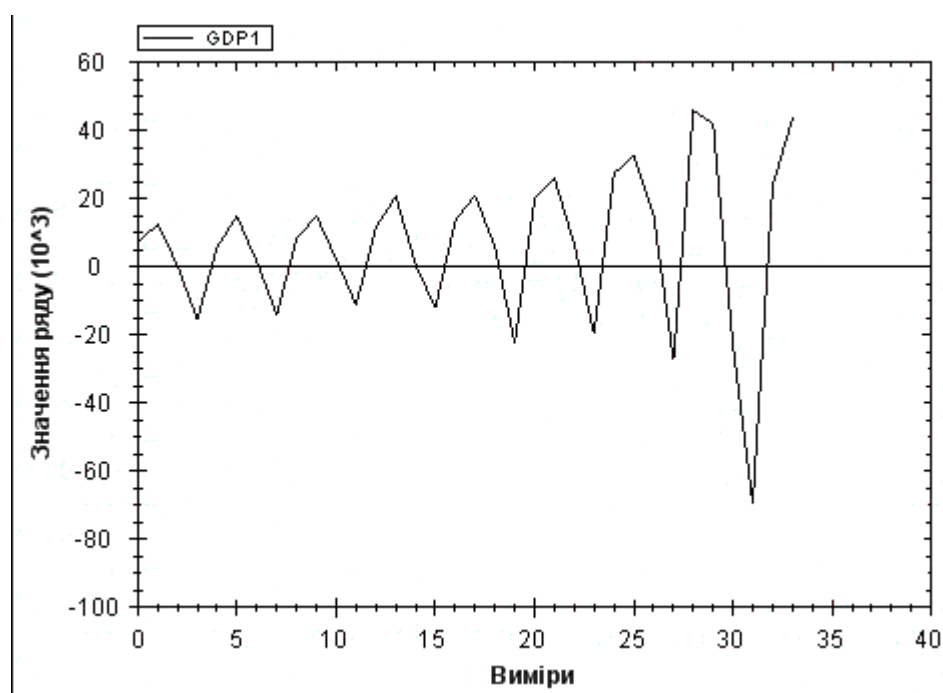


Рисунок 3.10 — Ряд ВВП України після взяття перших різниць

Після розрахунків АКФ та ЧАКФ (рис. 3.11, 3.12) для подальшого аналізу виберемо такі моделі АРКС:

$$1) y_1(k) = a_2 * y_1(k-2) + a_5 * y_1(k-5) + e(k) + b_3 * e(k-3) + b_4 * e(k-4) + b_5 * e(k-5)$$

$$2) y_1(k) = a_2 * y_1(k-2) + a_4 * y_1(k-4) + a_5 * y_1(k-5) + e(k) + b_1 * e(k-1)$$

де

$$y_1(k) = \text{ВВП}(k) - \text{ВВП}(k-1)$$

Статистичні характеристики		Статистичні характеристики	
Параметри якості моделі.		Параметри якості моделі.	
Кооф. детермінації	0,6528	Кооф. детермінації	0,9872
СКП	18602,8407	СКП	11781,0849
DW	2,2125	DW	0,8660
F-статистика	1,8798	F-статистика	230,6912
Кр. Тейла	0,3970	Кр. Тейла	0,2325

а) АРКС(5,5)

б) АРКС(5,1)

Рисунок 3.11 — Характеристики моделей



Рисунок 3.12 — АКФ та ЧАКФ ВВП після взяття перших різниць

Після обчислення характеристик вибраних моделей (Рис.4.8) остаточно зупиняємось на другому рівнянні, що краще описує наш процес:

$$y_1(k) = a_2 * y_1(k-2) + a_4 * y_1(k-4) + a_5 * y_1(k-5) + e(k) + b_1 * e(k-1)$$

де

$$y_1(k) = ВВП(k) - ВВП(k-1)$$

У табл. 3.3 та 3.4 представлено результати прогнозування ВВП на базі вибраної моделі АРКС(5,1).

Таблиця 3.3. Оцінка параметрів АРКС(5,1)

Коефіцієнт	Оцінка в СППР, метод МНК	Оцінка в СППР, метод РМНК
a_2	-0,3258	-0,3258
a_4	0,9875	0,9875
a_5	-0,1574	-0,1574
b_1	0,6493	0,6493

Таблиця 3.4. Результати прогнозування ВВП України

Вимір	Реальнезначення, млн грн	СППР			
		МНК		РМНК	
		Прогноз, млн грн	Похибка, млн грн	Прогноз, млн грн	Похибка, млн грн
2008-4	252 670	268 803	-16 133	268 803	-16 133
2009-1	183 217	210 871	-27 654	210 871	-27 654
2009-2	207 096	242 582	-35 486	242 570	-35 474
2009-3	250 614	265 792	-15 178	265 788	-15 174
2009-4		214 168		214 171	
Сер.абсолютна похибка			-18 890		-18 887

Вимір	Реальнезначення, млн грн	СППР			
		МНК		РМНК	
		Прогноз, млн грн	Похибка, млн грн	Прогноз, млн грн	Похибка, млн грн
Сер.абсолютна похибка, %			11,2%		11,2%

3.9. Дослідження та моделювання ВВП США

Дослідимо щоквартальні показники ВВП США, на основі значень яких побудуємо авторегресійну модель і зробимо однокроковий статичний прогноз на майбутнє.

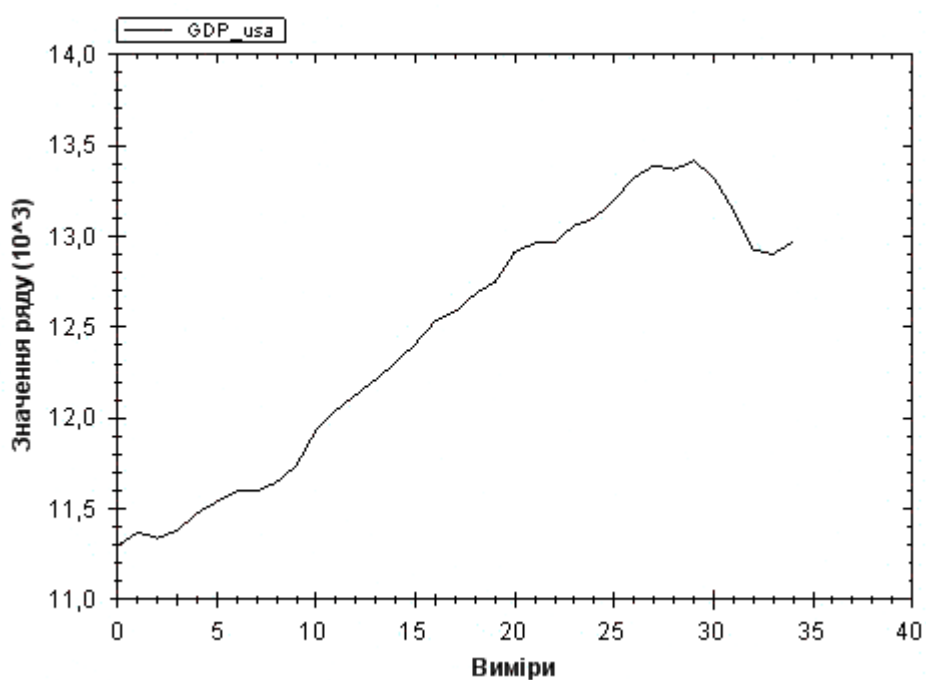


Рисунок 3.13 — Поквартальні показники ВВП США, починаючи із 2002 року, млрд дол.

Із графічного представлення даних можемо зробити припущення про наявність тренду (рис. 3.13). Провівши тест на нестационарність (рис. 3.14), переконуємося, що наша гіпотеза є вірною і перед подальшими дослідженнями та моделюванням варто виконати попередню обробку даних, щоб привести ряд до стаціонарного виду.

Тест Дікі-Фуллера. Ряд "GDP_usa"		
Тобч.	Ткр. при 5%	Ткр. при 1%
3,250	-1.94	-2.57

Рисунок 3.14 — ВВП США, результати тесту на не стаціонарність

Візьмемо перші різниці і отримаємо стаціонарний ряд, з яким будемо працювати в подальшому (рис. 3.15)

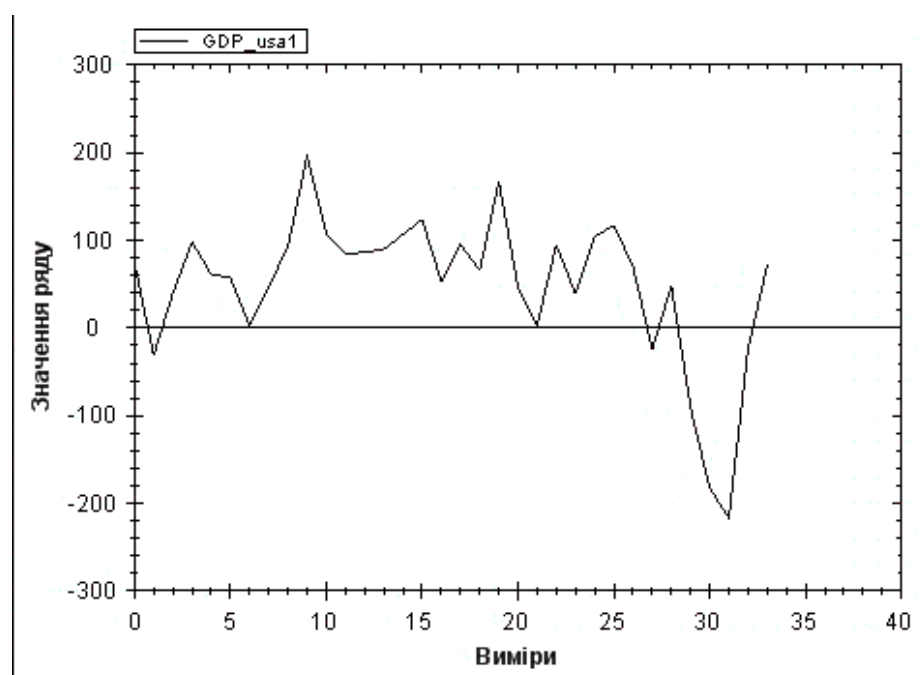


Рисунок 3.15 — Ряд ВВП США після взяття перших різниць

Після розрахунків АКФ та ЧАКФ (рис. 3.16) для подальшого аналізу виберемо модель АРКС(1,3). Проте її характеристики не зовсім задовільні, зокрема коефіцієнт детермінації (рис. 3.17). Тому розглянемо ще модель АРКС(2,3).

Отже, порівняємо такі моделі:

$$1) y_1(k) = a_1 * y_1(k-1) + e(k) + b_2 * e(k-2) + b_3 * e(k-3)$$

$$2) y_1(k) = a_1 * y_1(k-1) + a_2 * y_1(k-2) + e(k) + b_2 * e(k-2) + b_3 * e(k-3)$$

де

$$y_1(k) = ВВП(k) - ВВП(k-1)$$

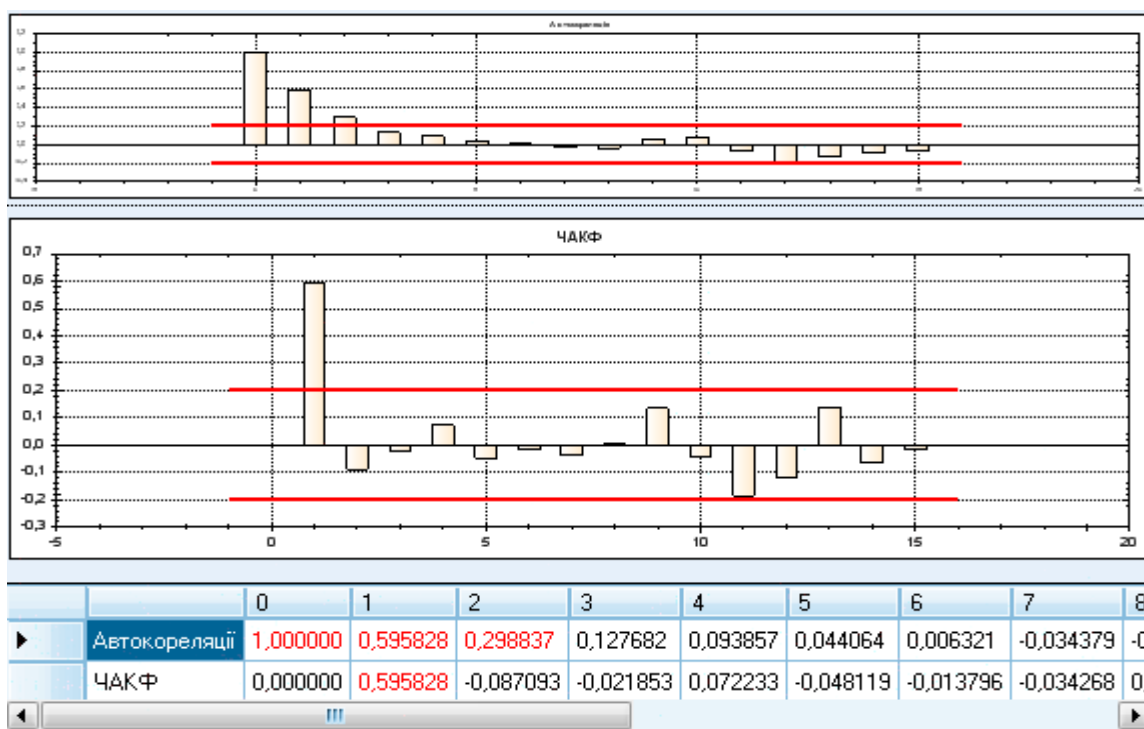


Рисунок 3.16 — АКФ та ЧАКФ для перших різниць

Параметри якості моделі.		Параметри якості моделі.	
Кооф. детермінації	0,6225	Кооф. детермінації	0,8410
СКП	70,4000	СКП	74,3238
DW	1,9777	DW	1,6577
F-статистика	11,5445	F-статистика	26,4378
Кр. Тейла	0,4039	Кр. Тейла	0,3916

а)АРКС(1,3) б)АРКС(2,3)

Рисунок 3.17 — АКФ та ЧАКФ для перших різниць

Після аналізу характеристик вибраних моделей остаточно зупиняємося на другому рівнянні, що краще описує наш процес:

$$y_1(k) = a_1 * y_1(k-1) + a_2 * y_1(k-2) + e(k) + b_2 * e(k-2) + b_3 * e(k-3)$$

де

$$y_1(k) = ВВП(k) - ВВП(k-1)$$

У табл. 3.5 та 3.6 представлено результати прогнозування ВВП на базі вибраної моделі АРКС(2,3).

Таблиця 3.5 — Результати прогнозування ВВП США

Вимір	Реальне значення, млрд дол.	СППР			
		МНК		РМНК	
		Прогноз, млрд дол.	Похибка, млрд дол.	Прогноз, млрд дол.	Похибка, млрд дол.
2008-4	13 141,9	13 269,6	-127,7	13 269,3	-127,4
2009-1	12 925,4	12 985,0	-59,6	12 985,1	-59,7

Вимір	Реальне значення, млрд дол.	СППР			
		МНК		РМНК	
		Прогноз, млрд дол.	Похибка, млрд дол.	Прогноз, млрд дол.	Похибка, млрд дол.
2009-2	12 901,5	12 721,3	180,2	12 721,4	180,1
2009-3	12 973,0	12 831,6	141,4	12 831,6	141,4
2009-4		13 020,0		13 020,0	
Сер.абсолютна похибка			26,9		26,9
Сер.абсолютна похибка, %			1,0%		1,0%

Таблиця 3.6 — Параметри моделі АРКС(2,3)

Коефіцієнт	Оцінка в СППР, метод МНК	Оцінка в СППР, метод РМНК
a_1	0,7388	0,7388
a_2	0,2417	0,2417
b_2	-0,3175	-0,3175
b_3	-0,3663	-0,3663

РОЗДІЛ 4. СТАРТАП-АНАЛІЗ ПРОЕКТУ

4.1. Інформаційна карта проекту

Опис інформації проекту виконано у таблиці 4.1.

Таблиця 4.1 — Опис інформації проекту

1. Назва проекту	Моделювання нестаціонарних процесів ціноутворення
2. Автори проекту	Чупрін Д.С.
3. Коротка анотація (не більше 1/3 сторінки)	Дана система призначена для моделювання та аналізу нестаціонарних процесів ціноутворення. На основі наявних історичних даних, система прогнозує кількість ціну на певний час у майбутньому
4. Термін реалізації проекту	6 місяців
5. Необхідні ресурси	Обладнання – комп'ютери, принтер, сканер, ноутбуки, планшети. Програмне забезпечення, операційна система, антивірусне обладнання. Електрика, газ, водопостачання, Інтернет. Фінансові ресурси – заробітна плата працівникам на 6 місяців роботи, гроші на оплату комунальних послуг, аренди, реклами тощо. Приміщення з усіма необхідними комунікаціями.
6. Опис проблеми, яку вирішує проект	Дана комплексна система дозволяє обчислити оцінки короткострокових прогнозів досліджуваних процесів та проаналізувати їх якість. На основі прогнозу відбувається прийняття рішення щодо подальших торгових операцій з досліджуваними акціями.
8. Головні цілі та завдання проекту	Основна мета проекту – отримання диплому автором цього проекту. Додаткові завдання – новий досвід,

	розробка комплексної системи, робота із реальними даними та створення комерційно успішного продукту.
9. Очікувані результати	
Система здатна будувати модель на основі даних про ціни акцій на певному відрізку часу, оцінювати короткострокові прогнози та аналізувати їх якість.	

4.2. Технологічний аудит ідеї проекту

Ідеї стартап-проекту описані у таблиці 4.2

Таблиця 4.2 — Опис ідеї стартап-проекту

Зміст ідеї	Напрямки застосування	Вигоди для користувача
Дана система дозволяє розв'язати проблему Обчислення оцінки короткострокових прогнозів досліджуванних процесів та аналізування їх якості. На основі прогнозу лице, що приймає рішення виконуватиме, відповідні до зроблених на основі прогнозу висногах, торгові операції	1. Робота із акціями на біржі.	Система дозволяє замовникам зменшити витрати часу та людських ресурсів на створення прогнозу, Покращити швидкість та якість прийняття рішень.
	2. Робота з трейденговими компаніями.	

Сильні, слабкі та нейтральні характеристики проекту описані у таблиці 4.3.

Таблиця 4.3 — Визначення сильних, слабких та нейтральних характеристик ідеї проекту

№ п/п	Техніко-економічні характеристики ідеї	(потенційні) товари/концепції конкурентів				W (слабка сторона)	N (нейтральна сторона)	S (сильна сторона)
		Мій проект	PhLab	Infosoftware Forecasting	Lokad			
1.	Точність прогнозування	Застосування кількох моделей в залежності від завдання	Власний алгоритм	Відсутнє прогнозування кількості дзвінків	Алгоритм Lokad			+
2.	Складання розкладу операторів	Наявне складання приблизного розкладу	Дана функція повністю відсутня	Дана функція повністю відсутня	Дана функція повністю відсутня			+
3.	Ризики невірного прогнозу	Існують, через велику кількість факторів.	Невідомо	Відсутні через відсутність прогнозу	Великий ризик невірного прогнозу		+	
4.	Доступність, зручність	Алгоритм, що виведе результати, на основі вхідних даних	Власний інтерфейс	Власний інтерфейс	Власний інтерфейс		+	

Технологічну здійсненність проекту описано у таблиці 4.4.

Таблиця 4.4 — Технологічна здійсненність ідеї проекту

Ідея проекту	Технології її реалізації	Наявність технологій	Доступність технологій
Створення системи короткострокових прогнозів	Використання статистичного пакету Eviews	Наявна	Доступні
Обрана технологія реалізації ідеї проекту: статистичний пакет Eviews			

4.3. Аналіз ринкових можливостей запуску стартап-проекту

Попередня характеристика потенційного ринку стартап-проекту описана у таблиці 4.5.

Таблиця 4.5 — Попередня характеристика потенційного ринку стартап-проекту

№ п/п	Показники стану ринку (найменування)	Характеристика
1	Кількість головних гравців, од	3
2	Загальний обсяг продаж, грн/ум.од	1000
3	Динаміка ринку (якісна оцінка)	Зростає
4	Наявність обмежень для входу (вказати характер обмежень)	Немає
5	Специфічні вимоги до стандартизації та сертифікації	Немає
6	Середня норма рентабельності в галузі (або по ринку), %	14%

Характеристику потенційних клієнтів стартап-проекту описано у таблиці 4.6 .

Таблиця 4.6 — Характеристика потенційних клієнтів стартап-проекту

№ п/п	Потреба, що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
1	Статистичні параметри роботи кол-центру, місячний прогноз	Аналітики, аналітичні відділи	Низька ціна, велика кількість статистичних даних.	Простота використання.
2	Створення якісного короткострокового прогнозу	Власники акцій, трейдери, трейдерськи компанії	Цікавить простота у використанні, низька ціна клієнська підтримка	Низька ціна, репутація

Фактори загроз описано у таблиці 4.7.

Таблиця 4.7 — Фактори загроз

№п/п	Фактор	Зміст загрози	Можлива реакція компанії
1.	Конкуренція	Програми для прогнозування ціноутворення будуть актуальні завжди, а тому не дивно, що з кожним роком, вони розвиваються і кількість їх збільшується.	Пришвидшити вихід програмного продукту
2.	Збут	Ускладнення збуту через цінову політику конкурентів	Розміщення додаткових рекламних банерів в Інтернеті

Фактори можливостей описано у таблиці 4.8.

Таблиця 4.8 — Фактори можливостей

№ п/п	Фактор	Зміст можливості	Можлива реакція компанії
1	Гнучкі ціни	Зменшення ціни товару задля збільшення попиту	Введення власних гнучких цін
2	Диференціація витрат	Зменшення витрат за рахунок їх перерозподілу	Зменшення витрат на додаткові, непрофільні задачі.

Ступеневий аналіз конкуренції на ринку описано у таблиці 4.9.

Таблиця 4.9 —Ступеневий аналіз конкуренції на ринку

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
1. Тип конкуренції досконала конкуренція	Багато систем/команд аналітиків	Розробити впізнаваний продукт, якість, що вирізнятиме нас від конкурентів
2. За рівнем конкурентної боротьби: міжнародний	На ринку присутні системи, розроблені за кордоном.	Розширення аудиторії, розширення списку мов, які підтримуються системою
3. За галузевою ознакою міжгалузева	Робота з акціями компаній різних галузей.	
4. Конкуренція за видами товарів: товарно-родова	Конкуренція між прогнозами інформаційних систем та команд аналітиків.	Збільшення функціоналу системи
5. За характером конкурентних переваг: нецінова	Різні способи прогнозування дають різну точність	Розробка кращих(точніших) алгоритмів
6. За інтенсивністю: марочна	Впізнаваний бренд надає великих переваг	Велику увагу приділити розвитку бренду

Аналіз конкуренції в галузі описано у таблиці 4.10.

Таблиця 4.10 — Аналіз конкуренції в галузі

Складові аналізу	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари-замінники
	Інші комплексні системи	Гнучкі ціни, розмір капіталовкладень	Змінні витрати постачальників, диференціація витрат	Контроль якості, система інформації	Ціна, лояльність споживачів
Висновки	Інтенсивна конкуренція	Є як можливості входження на ринок, так і нові потенційні конкуренти	Постачальники не диктують умови роботи на ринку	Клієнти не диктують умови роботи на ринку	Обмеження для роботи на ринку через товари-замінники

Обґрунтування факторів конкурентоспроможності описано у таблиці 4.11.

Таблиця 4.11 — Обґрунтування факторів конкурентоспроможності

№ п/п	Фактор конкурентоспроможності	Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим)
-------	-------------------------------	-----------------------------------------------------------------------------------------------------

2	Якість	Висока якість прогнозу, велика кількість допоміжних статистичних даних
3	Обслуговування	Переважає кількість клієнтів працюють з короткостроковими прогнозами, а тому маємо велику цільову аудиторію.

Порівняльний аналіз сильних та слабких сторін «назва проекту» описано у таблиці 4.12.

Таблиця 4.12 — Порівняльний аналіз сильних та слабких сторін «назва проекту»

№ п/ п	Фактор конкурентоспроможності	Бали 1-20	Рейтинг товарів-конкурентів						
			-3	-2	-1	0	+1	+2	+3
1	Багатофункціональність	15		2			3		1
2	Якість	20	3	2	1		2		
3	Обслуговування	5					3	1	2

SWOT-аналіз стартап-проекту описано у таблиці 4.13.

Таблиця 4.13 — SWOT-аналіз стартап-проекту

Сильні сторони: Висока якість прогнозу Якість Багатофункціональність	Слабкі сторони: Інтерфейс користувача Немає налагодженої клієнської бази
Можливості: Попит	Загрози: Конкуренція

Зміна рівня доходів населення	Збут
Вдосконалення системи	

Альтернативи ринкового впровадження стартап-проекту описано у таблиці 4.14 .

Таблиця 4.14 — Альтернативи ринкового впровадження стартап-проекту

№ п/п	Альтернатива (орієнтовний комплекс заходів) ринкової поведінки	Ймовірність отримання ресурсів	Строки реалізації
1	Швидкий вихід на ринок із «сирим» продуктом, можливі проблеми із точністю прогнозу та універсальністю	30%	3 місяці
2	Поступовий вихід з готовим, відлагодженим продуктом. Висока якість та конкурентоспроможна ціна	70%	6 місяців

4.4. Розроблення ринкової стратегії проекту

Вибір цільових груп потенційних споживачів описано у таблиці 4.15.

Таблиця 4.15 — Вибір цільових груп потенційних споживачів

№ п/п	Опис профілю цільової групи потенційних клієнтів	Готовність споживачів сприйняти продукт	Орієнтовний попит в межах цільової групи (сегменту)	Інтенсивність конкуренції в сегменті	Простота входу у сегмент
1	Окремі аналітики та аналітичні відділи невеликих компаній	Висока готовність	60%	Висока	Середня
2	Середні трейдерські компанії	Висока	35%	Середня	Середня
3	Великі гравці	Дуже низька	5%	Середня	Низька
Які цільові групи обрано: 2, 3					

Визначення базової стратегії розвитку описано у таюл 4.16

Таблиця 4.16 — Визначення базової стратегії розвитку

№ п/п	Обрана альтернатива розвитку проекту	Стратегія охоплення ринку	Ключові конкурентоспроможні позиції відповідно до обраної альтернативи	Базова стратегія розвитку*
1	2 та 3	Стратегія диференційованого маркетингу	Висока універсальність, багатогалузевість, висока якість, ціна	Стратегія лідерства в простоті та доступності подання інформації до ЛПР

Визначення базової стратегії конкурентної поведінки представлено у таблиці 4.17.

Таблиця 4.17 — Визначення базової стратегії конкурентної поведінки

Чи є проект «першопрохідцем» на ринку?	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристики товару конкурента, і які?	Стратегія конкурентної поведінки*
Ні	Так	Ні	Стратегія виклику лідера

Визначення стратегії позиціонування описано у таблиці 4.18.

Таблиця 4.18 — Визначення стратегії позиціонування

Вимоги до товару цільової аудиторії	Базова стратегія розвитку	Ключові конкурентоспроможні позиції власного стартап-проекту	Вибір асоціацій, які мають сформувати комплексну позицію власного проекту (три ключових)
Якість, точність, простота у використанні	Стратегія лідерства по витратах	Якість прогнозу, універсальність, велика кількість статистичної інформації	По іміджу Позиціонування на низькій ціні позиціонування за сферою застосування

4.5. Розроблення маркетингової програми стартап-проекту

Визначення ключових переваг концепції потенційного товару описано у таблиці 4.19.

Таблиця 4.19 — Визначення ключових переваг концепції потенційного товару

№ п/п	Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
1	Статистичні данні щодо цін на певному проміжку часу	Розгорнута статистика, яка охоплює багато коефіцієнтів, що мають значення при прийнятті	Дана статистика має бути максимально розгорнутою та зрозумілою.

№ п/п	Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
		рішення.	
2	Створення якісного та точного прогнозу	Точний прогноз – основа успіху при торгових операціях	Розробка короткострокових прогнозів, використання різних методів для покращення точності прогнозу

Опис рівнів сутності та складових товару описано у таблиці 4.20.

Таблиця 4.20 — Опис рівнів сутності та складових товару

Рівні товару	Сутність та складові		
I. Товар за задумом	Комплексна система обслуговування прийняття рішень щодо акцій		
II. Товар у реальному виконанні	Властивості/характеристики	М/Нм	Вр/Тх /Тл/Е/Ор
	Програмний продукт – складна, комплексна система		
	Якість – тестування сторонніми фірмами		
	Пакування – зручний інсталятор		
	Марка: назва організації-розробника + назва товару		
III. Товар із підкріпленням	До продажу: доставка, гнучкі умови оплати		
	Після продажу: гарантія якості		
За рахунок чого потенційний товар буде захищено від копіювання: Система авторського права			

Визначення меж встановлення ціни описано у таблиці 4.21.

Таблиця 4.21 — Визначення меж встановлення ціни

Рівень цін на товари-замінники	Рівень цін на товари-аналоги	Рівень доходів цільової групи споживачів	Верхня та нижня межі встановлення ціни на товар/послугу
1000	1500	300000	800-900

Формування системи збуту описано у таблиці 4.22.

Таблиця 4.22 — Формування системи збуту

Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту
Канал нульового рівня	Доставка товару	0	+

Висновки до розділу 4

Розглянуто стартап-ідею проекту.

З'ясовано, що є можливість ринкової комерціалізації проекту (наявні попит, динаміка ринку, рентабельність роботи на ринку);

З'ясовано, що є перспективи впровадження з огляду на потенційні групи клієнтів, бар'єри входження, стан конкуренції. Конкурентоспроможність проекту є високою;

Подальша імплементація проекту є можливою.

ВИСНОВКИ

Розглянуто і проаналізовано два найбільш вживані, з існуючих на даний момент, продуктів для статистичної обробки даних. Проведений огляд деяких відомих методів прогнозування часових рядів, а також розглянуті підходи до тестування ряду на нелінійність та на наявність тренду.

Побудовано модель для нестационарних процесів у ціноутворенні акцій компанії Intel. В проекті представлені результати прогнозування вибраних нестационарних часових рядів.

Розглянуто шляхи можливого подальшого вдосконалення системи.

ПЕРЕЛІК ПОСИЛАНЬ

1. Ставицький А.В. Навчально-методичний комплекс з курсів „Прогнозування” та „Фінансове прогнозування” / Ставицький А.В. — К., 2006. — 107 с.
2. Бідюк П.И. Структурный анализ методик построения регрессионных моделей по временным рядам наблюдений URL / Бідюк П.И., Зворыгина Т.Ф. — Режим доступа: <http://masters/donntu.edu.ua/2006/fvti/sergienko/library/art03.htm>
3. Бідюк П.И. Система підтримки прийняття рішень для прогнозування нестационарних процесів /Бідюк П.И., Демківський Є.О. //Наукові праці Миколаївського державного технічного університету. — 2008. — Випуск 77. — С. 137—159.
4. Tsay R.S. Analysis of financial time series / Tsay R.S. — NewYork: John Wiley&Sons, Ltd, 2002. — 455 p.
5. Бідюк П.И. Аналіз часових рядів / Бідюк П.И., Романенко В.Д., Тимошук О.Л. — К.: НТУУ «КПІ», Політехніка, 2013. — 600 с.
6. Бідюк П.И. Економетричний аналіз часових рядів: Конспект лекцій / Бідюк П.И. — К.: НТУУ «КПІ», 2007. — 250 с.
7. Xekalaki E. ARCH models for financial applications / Xekalaki E., Degiannakis S. — New York: John Wией & Sons Ltd., 2010. — 550 p.
8. Engle F.R. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation / Engle F.R. // Econometrica. — 1982, vol. 50, no. 4. — P. 987–1007.
9. Engle R.F. Modeling the persistence of conditional variance / Engle R.F., Bollerslev T. //Econometric Reviews. — 1986. Vol. 5. P. 1—50.
10. Holsapple C.W. Decision Support Systems / Holsapple C.W., Whinston A.B. — New York: West publishing company. 1996. 860 p.

11. Bollerslev T.B. Generalized autoregressive conditional heteroscedasticity /T.B. Bollerslev // Journal of Econometrics. — 1986. — № 3. — pp. 307—327.
12. Bauwens L. Multivariate GARCH models: a survey/ Bauwens L. //Journal of Applied Econometrics. — 2006. vol. 21. — P. 79—100.
13. Nelson D.B. Conditional Heteroskedasticity in Asset Returns: A New Approach /D.B.Nelson //Econometrica. — 1991. — № 2. — P. 347—370.